

## **A thermostable in vitro polymerase complex for template-dependent elongation of nucleic acids in amplification or reverse transcription methods**

**Patent number:** DE19840771  
**Publication date:** 2000-02-10  
**Inventor:** VOSS HARTMUT [DE]; MOECKEL GERD [DE]; KOBER INGO [DE]; KILGER CHRISTIAN [DE]  
**Applicant:** LION BIOSCIENCE AG [DE]  
**Classification:**  
- **international:** C12N9/12; C07K19/00; C12P19/34; C12N15/62; C12N15/63; C12Q1/68  
- **european:** C07K14/195; C12N9/12B7B7; C12N9/12B7B49  
**Application number:** DE19981040771 19980907  
**Priority number(s):** DE19981040771 19980907; DE19981035653 19980806

### **Abstract of DE19840771**

A thermostable in vitro complex for template-dependent elongation of nucleic acids comprises a thermostable sliding clamp protein, which is connected with an elongation protein that shows thermostable polymerase activity is new. Independent claims are also included for the following: (1) a thermostable accessory in vitro complex characterized in that it contains a sliding clamp protein and a coupling protein, as defined above; (2) a recombinant DNA sequence, which encodes a thermostable in vitro complex or a thermostable accessory complex; (3) a vector that contains recombinant DNA encoding a sliding clamp protein and a coupling protein and/or an elongation protein; (4) a host cell transformed with one or more vectors of (3); (5) a method for the production of a thermostable in vitro complex or thermostable in vitro accessory complex as above; (6) a method for template-dependent elongation of nucleic acids; (7) a method of marking nucleic acids through generating a single break in the phosphodiester bond of the nucleic acid chain and substituting a nucleotide at the break site with a marked nucleotide with the help of a polymerase, where the polymerase is a thermostable in vitro complex as above; and (8) a reagent kit for elongation and/or amplification and/or reverse-transcription and/or sequencing and/or marking of nucleic acids.

---

Data supplied from the esp@cenet database - Worldwide

**THIS PAGE BLANK (USPTO)**

## **Laid Open Publication**

### **DE 198 40 771 A 1**

---

#### **Description**

The present invention relates to a thermostable in vitro complex for template-dependent elongation of nucleic acids, a thermostable, prokaryotic, accessory in vitro complex, as well as DNA-sequences and vectors for encoding it. The invention further relates to the use of the inventive complex in methods of template-dependent elongation of nucleic acids, such as PCR reactions or those of DNA sequencing involving in vitro template-dependent DNA strand synthesis. Finally, the present invention relates to reagent kits for implementing the inventive method.

DNA polymerases are members of a group of enzymes that utilize single stranded DNA as templates in the synthesis of a complementary DNA strand. These enzymes play an important role in nucleic acid metabolism, including the processes of DNA replication, repair and recombination. DNA polymerases have been identified in all cellular organisms, ranging from bacterial to human cells, in many viruses and in bacteriophages (Kornberg, A. & Baker, T.A. (1991) DNA Replication WH Freeman, New York, NY). Generally, archaebacteria and eubacteria are grouped together in the group of prokaryotes, organisms lacking a true cell nucleus, and are contrasted with eukaryotes, organisms having a true cell nucleus. Common to many polymerases from a wide variety of organisms are often similarities in amino acid sequence and structure (Wang, J., Sattar, A.K.M.A; Wang, C.C., Karam, J.D., Konigsberg, W.H. & Steitz, T.A. (1997) "Crystal Structure of pol a family replication DNA polymerase from bacteriophage RB69", Cell 89, 1087-1099). Organisms, such as human organisms, possess a plurality of DNA-dependent polymerases, not all of which however are responsible for DNA replication. Several also effect DNA repair. Replicating DNA polymerase consists for the most part of protein complexes having multiple subunits which replicate the chromosomes of cellular organisms and viruses. A general characteristics of such replicating polymerases is a generally high degree of processivity, that is, their ability to polymerize thousands of nucleotides without disassociating from a DNA template (Kornberg, A & Baker, T.A. (1991) DNA Replication. WH Freeman, New York, NY).

Until recently, the only well understood highly processive replication mechanisms in cells were those used by cellular replicases (a protein complex exhibiting polymerase activity), in addition to the replication apparatus of the bacteriophages T4 or T7. The apparatus underlying this mechanism contains a protein having a ring-like structure, a "sliding" clamp which encircles the DNA, and binds the catalytic polymerase unit – the "elongation" protein to the DNA (Stukenberg, P.T., Studwell-Vaughan, P.S. & O'Donnell, M. (1991) "Mechanism of the  $\beta$ -clamp of DNA polymerase III holoenzyme", J. Biol. Chem. 266, 11328-11334; Kuriyan, J. & O'Donnell, M. (1993) "Sliding clamps of DNA polymerases" J. Mol. Biol. 234, 915-925. Frequently, the sliding clamp is bonded via one or more additional proteins, "coupling proteins"

**THIS PAGE BLANK (USPTO)**

to the elongation protein. The 3-dimensional structure of various sliding clamp proteins has already been determined:

- that of the eukaryotic proliferating cell nuclear antigen (PCNA) (Krishna, T.S.R., Kong, X.-P., Gary, S., Burgers, P.M. & Kuriyan, J. (1994) "Crystal structure of the eukaryotic DNA polymerase processivity factor PCNA", *Cell* 79, 1233-1243; Gulbis, J.M., Kelman, Z., Hurwitz, J., O'Donnell, M. & Kuriyan, J. (1996) "Structure of the C-terminal region of p21 WAF1/CIP1 complexed with human PCNA", *Cell* 87, 297-306),
- that of the  $\beta$  subunit of polymerase III of the eubacterium *Escherichia coli*, (Kong, X.-P., Onrust, R., O'Donnell, M. & Kuriyan, J. (1992) "Three dimensional structure of the  $\beta$  subunit of *Escherichia coli* DNA polymerase III holoenzyme; a sliding DNA clamp", *Cell* 69, 425-437)
- and that of the bacteriophages T4 Gen45 proteins (Kelman, Zvi, Hurwitz, J. O'Donnell, Mike (1998) *Structure* 6, 121-125).

Such sliding clamps are very similar in their overall structure: Superimposed, the images of the overall protein structure of PCNA, the  $\beta$  subunit, and of the gp45 rings are congruent (Kelman, Z. & O'Donnell, M. (1995) "Structural and functional similarities of prokaryotic and eukaryotic sliding clamps", *Nucleic Acids Res.* 23, 3613-3620). Each ring is comparably dimensioned with a central opening large enough to encircle duplex-DNA, that is, a double strand of DNA composed of two complementary DNA strands.

In vivo, the sliding clamp itself is unable to position itself around the DNA; rather it has to be fixed around the DNA by means of a clamp loader. Such a clamp loader is a protein complex which in prokaryotes and eukaryotes consists of a plurality of subunits, and which is called  $\gamma$ -complex in conjunction with the eubacterium *Escherichia coli*, and replication factor C (RF-C) in conjunction with humans (Kelman, Z. & O'Donnell, M. (1994) "DNA replication – enzymology and mechanisms", *Curr. Opin. Genet. Dev.* 4, 185-195). The sliding clamp recognizes the 3'-end of the single strand duplex (primer-template) and positions the sliding clamp in the presence of ATP around the DNA. The sliding clamp which encircles the DNA interacts with polymerases, thereby ensuring rapid and processive DNA synthesis.

In the case of bacteriophage T7, the same goal, i.e. processive DNA synthesis, is achieved by means of a structurally different protein complex. The phage expresses its own catalytic polymerase, the T7 polymerase, the product of gene 5, which enters into a bond with a protein derived from the host *Escherichia coli*, thioredoxin and, as a replicase, enables highly processive DNA replication (*Proc Natl Acad Sci USA* October 15, 1992; 89(2):9774-9778 "Genetic analysis of the interaction between bacteriophage T7 DNA polymerase and *Escherichia coli* thioredoxin", Himawan J.S., Richardson C.C.). This also leads to clamp formation; although the structure of said clamp is not the same as, for example, in the case of eukaryotic PCNA.

It is often necessary -- as for example in the case of human polymerase  $\delta$  --, for proteins (coupling proteins) to create a link between the catalytically active portion of the polymerase and the processivity factor (sliding clamp). In humans, this is the small subunit of the  $\delta$  polymerase (Zhang, S.-J., Zeng, X.-R., Zhang, P., Toomey, N.L., Chuang, R.Y., Chang, L.-S., and Lee,

THIS PAGE BLANK (USPTO)

M.Y.W.T. (1994), "A conserved region in the amino terminus of DNA polymerase  $\delta$  is involved in proliferating cell nuclear antigen binding", *J. Biol. Chem.* 270, 7988-7992. In the case of T7 polymerase, however, the processivity factor directly binds the catalytically active unit of the polymerase.

DNA polymerases are characterized by two properties, among others, their rates of elongation, that is, the number of nucleotides they are able to incorporate per second into a growing strand of DNA, and their dissociation constants. If, after each incorporation step, a polymerase again dissociates from the strand one of the nucleotides in the growing chain (that is, one elongation step occurs per bonding event), processivity then has a value of 1 and the polymerase is non-processive. If the polymerase remains bonded to the strand during repeated nucleic acid incorporation, then the replication modus is termed processive and may achieve a value of several thousand (see also: *Methods in Enzymology*, Volume 262, DNA Replication, Edited by J.L. Campbell, Academic Press 1995, pp. 270-280).

For most in vitro applications, such as PCR or sequencing processes, processivity is a desirable characteristic, but one which to date is possessed by thermostable enzymes used in such reactions to only a small degree. In contrast, the temperature sensitive T7 polymerase associated with thioredoxin exhibits a processivity of several thousand nucleotides. By comparison, the *Thermus thermophilus* or *aquaticus*-derived thermostable DNA polymerases have a processivity of approximately 50 nucleotides (*Biochim Biophys Acta* Nov. 7, 1995; 1264(2):243-248, "Inactivation of the 5'-3' exonuclease of *Thermus aquaticus* DNA polymerase", Markens L.S., Bryan, S.K., Mokses R.E.).

U.S. Patents 4,683,195, 4,800,195 and 4,683,202 describe the application of such thermostable DNA polymerases in polymerase chain reactions (PCR). In a PCR, a DNA is newly synthesized using primers, templates, nucleotides a DNA polymerase of a corresponding buffer and under suitable reaction conditions. During the process the double-stranded target sequence is most commonly fused, two oligonucleotides are hybridized thereto and the complementary sequence is then synthesized by the polymerase on the template through incorporation of nucleotides. The extension product of each primer functions as a template for the next cycle. In PCR of this type, it is preferable to use a thermostable polymerase which survives the cyclical, thermal fusion of the DNA strands. Thus, Taq DNA polymerase is frequently used (U.S. Patent 4,965,188). However, as described above, the processivity of Taq DNA is relatively small as compared to T7 polymerase.

DNA polymerases are also used in DNA sequence determination (Sanger et al., *Proc. Natl. Acad. Sci., USA* 74:5463-5467 (1997)). Per Sanger, in sequence determination a T7 DNA polymerase is frequently used (Tabor, S. and Richardson, C.C., *Proc. Natl. Acad. Sci., USA* 86:4076-4080 (1989)). Subsequently, the cycle-sequencing method was developed (Murray, V. (1989) *Nucleic Acids Res.* 17, 8889), which does not require a single stranded template and which permits initiation of the sequence reaction with relatively small amounts of template. Polymerases utilized in this method may include, for example, the aforementioned Taq polymerase (U.S. Patent 5,075,216) or *Thermotoga neapolitana*-derived polymerase (WO96/10640), or other thermostable polymerases. Newer methods combine the exponential amplification and sequencing of a DNA fragment in one step, allowing for direct sequencing of

**THIS PAGE BLANK (USPTO)**



genomic DNA. One such method, the so-called DEXAS-method (Nucleic Acid Res., May 15, 1997; 25(10): 2032-2034, "Direct DNA sequence determination from total genomic DNA", Kilger C., Pääbo, S., Biol Chem, Feb. 1997; 378(2):99-105, "Direct exponential amplification and sequencing DEXAS) of genomic DNA", Kilger, C., Pääbo, S, and DE 196 53 439.9, as well as DE 196 53 494.1) employs a polymerase with reduced discrimination capacity vis-à-vis dideoxynucleotides (ddNTPs) as compared with deoxynucleotides (dNTPs), further a reaction buffer, two primers which are preferably not equimolar, and the aforementioned nucleotides, whereby in multiple cycles a complete, sequence-specific DNA ladder of a fragment is then obtained which is encompassed by the primers. In a further modification of this method a polymerase mix is used, in which one of the two polymerases discriminates between ddNTPs and dNTPs, while the other displays a reduced discrimination capacity (Nucleic Acids Res., May 15, 1997; 25(10):2032-2034, "Direct DNA sequence determination from total genomic DNA", Kilger C., Pääbo, S).

DNA polymerases are also used in reverse transcription of RNA into DNA. Here, RNA serves as a template and the polymerase synthesizes a complementary DNA strand, using, for example, the thermostable DNA polymerase derived from the *Thermus thermophilus* (Tth) organism (U.S. Patent 5,322,770).

It may also be desirable for the polymerase to exhibit "proof-reading" activity, that is a 3'-5' exonuclease activity. This characteristic is particularly desirable when the product being synthesized is to be generated at a low error rate during nucleotide incorporation.

The aforementioned enzymes that are commonly used in PCR-reactions, are not for the most part members of the group of actual replication enzymes, but rather are mostly enzymes that are assumed to participate in DNA repair, which accounts for their low degree of processivity.

It is thus an object of the present invention to combine several of the aforementioned characteristics of polymerases, in particular high processivity and thermostability, for use in *in vitro* reactions.

This objective has been achieved in accordance with the present invention by the preparation of a thermostable *in vitro*-complex for template-dependent elongation of nucleic acids, comprising a thermostable sliding clamp protein that is connected to an elongation protein exhibiting thermostable polymerase activity. The complex may be utilized in *in vitro* reactions, such as PCR-reactions in which it exhibits high processivity. It is also advantageous if the complex exhibits a low error rate during nucleotide incorporation, that is, it possesses enhanced fidelity. Hence, it is possible to use the complex in conjunction with elongation, amplification and sequencing of nucleic acids. The complex may preferably be utilized in Standard-PCR-reactions.

To ensure the suitability of such a complex in Standard-PCR-reactions, it must be easy to manipulate. Thus, the reaction apparatus of the simian virus 40 has been constituted *in vitro* (Wage, S. & Stillman, B. (1994) *Nature*, vol. 369:207-221), but such a reaction apparatus is unsuitable for Standard-PCR reactions because among other things, it is not thermostable.

THIS PAGE BLANK (USPTO)

Hence, the object of the present invention is the preparation of a thermostable in vitro-complex exhibiting polymerase activity that is useable in Standard-PCR reactions and which displays high processivity.

In particular, the object of the present invention is a thermostable prokaryotic, in vitro complex for elongation of nucleic acids, which comprises a thermostable sliding clamp protein that encircles entirely or in part the complementary nucleic acid strand, and a thermostable protein exhibiting polymerase activity, wherein said protein or protein complex is coupled with the sliding clamp protein.

Within the scope of the present invention the concept of an elongation protein exhibiting polymerase activity also encompasses protein complexes that exhibit polymerase activity, or subunits of such complexes which carry the polymerase activity.

"Thermostable" as used in the present invention means that the accessory complex with high processivity incorporates nucleotides in growing nucleic acid chains, under both low and high temperatures that occur in PCR or other reactions, such as for example DNA sequencing.

As a rule, PCR consists, for example, of the steps of denaturing (70° to 98°C), annealing (40° to 78°C) and DNA strand synthesis (60° to 76°C). Therefore, the complex must be functional at least between about 60° and about 70°C, in particular between 60° to 76°C, and especially preferably between 40° to 98°C. No irreversible denaturing of the complex or individual components may occur over the course of the entire reaction that will stop or inhibit the elongation reaction.

The sliding clamp protein and the elongation protein exhibiting polymerase activity may be coupled to one another by covalent or non-covalent bonding. In a preferred embodiment the sliding clamp protein and the elongation protein are connected by means of a coupling protein.

In another preferred embodiment of the present invention the associated proteins within the inventive thermostable complex derive from archaeobacteria, though within the scope of the present invention it is also feasible that the associated proteins may derive from eubacteria. An object of the present invention is further a thermostable in vitro complex according to the present invention, in which the associated proteins are derived in part from archaeobacteria and in part from eubacteria.

Within the meaning of the present invention, the notion prokaryotic protein encompasses both proteins derived from archaeobacteria and proteins derived from eubacteria. The replication apparatus in archaea is known to be similar to the eukaryotic replication apparatus, even though the genomic organization in eukaryotes and archae are complete different, and the cellular structure of eubacteria resembles that of archaea (Edgell, D.R. and Doolittle, W.F. (1997), "Archaea and the origin(s) of DNA replication proteins", *Cell* 89:995-998.

Thus, a preferred embodiment of the present invention is a thermostable prokaryotic in vitro complex in which a protein complex exhibiting polymerase activity is present which consists of a coupling protein and one or more elongation proteins exhibiting polymerase

**THIS PAGE BLANK (USPTO)**

activity. Thus, a thermostable prokaryotic in vitro complex according to the present invention is preferred in which the sliding clamp protein has a ring-shaped structure that encircles entirely or in part the complementary nucleic acid strands.

### The Sliding Clamp Protein

The following explanations are intended to assist in better understanding the function and potential manifestations of the sliding clamp protein.

The sliding clamp protein serves the function of binding the polymerase activity to the DNA. Either the sliding clamp protein itself encircles the DNA completely or in part, or a clamp is formed by association to the protein exhibiting polymerase activity or to a protein complex exhibiting polymerase activity, or to a subunit thereof. In any case, the level of processivity is significantly increased, at least by a factor of  $1 \frac{1}{2}$  as a result of the clamp formation.

That means that the in vitro complex according to the present invention exhibits at least  $1 \frac{1}{2}$  times as much processivity as compared to the elongation protein alone, or as compared to a protein complex exhibiting polymerase activity without a sliding clamp, or a subunit thereof.

Functioning slide clamps may be, for example, human genome-derived homologs of the "Proliferating Cell Nuclear Antigen"-protein complex, or homologs also of the ring-shaped, E-coli-derived " $\beta$ -clamp"-protein complex, which are derived from thermostable organisms and thus are thermostable, or from non-thermostable organisms, which are subsequently rendered thermostable by altering the amino acid sequence (Eijsink V.G., van der Zee J.R., van den Burg B., Vriend G. Venema, G. FEBS Lett, April 22, 1991, 282(1):13-16, Improving the thermostability of the neutral protease of *Bacillus stearothermophilus* by replacing a buried asparagine by leucine", Bertus Van den Burg, Gert Vriend, Oene R. Veltman, Geard Venema, and Vincent G. H. Eijsink, "Engineering an enzyme to resist boiling", PNAS 1998, 95:2056-2060). The sliding clamp may be constructed of multiple components. The sliding clamp identified in the human genome consists of three PCNA-protein components (SEQ ID NO: 11) (homotrimers), the sliding clamp identified in the E. coli genome consists of two components (SEQ ID NO: 35) (homodimers).

Within the meaning of the present invention a slide clamp is understood to be in particular any protein that possesses the functional property of polymerase activity enhancement or serves to lower the error rate. In addition, the sliding clamp may exhibit a ring-shaped, three-dimensional structure, or form a ring-shaped, three-dimensional structure by coupling to another protein, which enables it to encircle, either entirely or in part, single and double-stranded DNA.

A sliding clamp within the meaning of the present invention is understood to be a protein which

1. exhibits in a sequence alignment to the human (eukaryotes) PCNA amino acid sequence (SEQ ID NO: 11) having a length of at least 100 amino acids at least a 20% sequence identity, or

**THIS PAGE BLANK (USPTO)**

2. exhibits in a sequence alignment to the bacterial  $\beta$ -clamp sequence derived from *E. coli* (eubacteria) (SEQ ID NO: 35) having a length of at least 100 amino acids at least a 20% sequence identity, or

3. exhibits in a sequence alignment to the amino acid sequence of the PCNA homologs derived from *Archaeoglobus Fulgidus* (Archaea) (SEQ ID NO: 12) having a length of at least 100 amino acids at least a 20% sequence identity.

The sliding clamp protein according to the present invention may have one or more of the aforementioned features.

The sequence identities cited in **Fig 1** were determined with the help of the BLAST Algorithm according to Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J., *J. Mol. Biol.* 215:403-410 (1990) and are discussed in detail below.

Sliding clamps within the meaning of the present invention are also understood to be proteins that contain one or both of the following consensus sequences and which deviate from these sequences at no more than four positions (**Abb.4**):

#### Region 1

(SEQ ID NO: 39)

[G A V L I M P F W]-D-X-X-X-[G A V L I M P F W]-X-X-[G A V L I M P F W]-X-[G A V L I M P F W]-X-[G A V L I M P F W]-X-X-X-X-F-X-X-Y-X-X-D and/or

#### Region 2

(SEQ ID NO: 40)

[G A V L I M P F W]-X(3)-L-A-P-[K R H D E]-[G A V L I M P F W]-E.

The amino acids are designed here using the standard IUPAC single letter nomenclature and cited in accordance with the Prosite Pattern description standard, and where the following amino acid groups are frequently lumped together:

G, A, V, L, I, M, P, F or W (amino acids with non-polar side chains)

S, T, N, Q, Y, or C (amino acids with non-charged polar side chains)

K, R, H, D or E (amino acids with charged and polar side chains)

In addition, X stands for any arbitrary amino acid, insertion or deletion in the sequence listing.

Also generated from the multiple alignment of human PCNA homologs shown in **Fig.12** was a Hidden Markov Model. Thus, a sliding clamp within the meaning of the present invention is understood to be in particular any protein that with a Hidden Markov Model so generated has a score of greater than 20 (**Fig. 12**). The Hidden Markov Models and corresponding scores were calculated using the hmmfs program (Version 1.8.4, July 1997) from the HMMER packet (HMMER Protein and DNA Hidden Markov Models (Version 1.8) by Sean Eddy, Dept. of Genetics, Washington University School of Medicine, St. Louis, USA).

THIS PAGE BLANK (USPTO)



A Hidden Markov Model was generated from the multiple alignment of *E. coli*  $\beta$ -clamp homologs shown in **Fig. 13**. Thus, a sliding clamp within the meaning of the present invention is understood to be in particular any protein that with a Hidden Markov Model so generated has a score of greater than 25 (**Fig. 13**).

The sliding clamp may be constructed of multiple components that are fixedly bound to one another by a characteristic bond to thereby form a stable ring-shaped molecular complex which cannot readily dissociate from the DNA. This makes possible a fixed but non-covalent bond to the DNA, but which does not inhibit free displaceability along the latter. In addition, the processivity-enhancing sliding clamp proteins possess characteristic local molecular properties in the region of interaction with the DNA, which improve free displaceability, and which may then be supported via water molecules stored in this region.

A preferred embodiment of the present invention is further in particular a thermostable prokaryotic in vitro complex, wherein the sliding clamp protein is one of the following: *Archaeoglobus Fulgidus*-derived AF0335, *Methanococcus Jannashcui*-derived MJ0247, *Pyrococcus Horikoschii*-derived PHLA008, *Methanobacterium Thermoautotrophicus*-derived MTH1312, as well as *Aquifex Aeolicus*-derived AE000761\_7.

In particular, the subject matter of the present invention encompasses thermostable prokaryotic in vitro complexes, in which the sliding clamp protein has an amino acid in accordance with SEQ ID NO: 11, 12, 13, 14, 15 and 36 (*Aquifex Aeolicus*).

#### Sliding Clamp Loader

In addition, a preferred embodiment is understood to be one in which part of the complex according to the present invention is a sliding clamp loader, which assembles the components of the sliding clamp around the continuous DNA strand, or removes them when the reaction is completed. Said sliding clamp loader is preferably associated with the in vitro complex of the present invention.

In humans the sliding clamp loader consists of five subunits, 4 small (sliding clamp loader 1) and one large subunit (sliding clamp loader 2). Functional sliding clamp loaders in accordance with the present invention are, for example, one or more prokaryotic homologs of "Replication"-Factor C-protein complex identified in humans (sliding clamp loader 1: SEQ ID NO: 1, 32, 33, 34 and sliding clamp loader 2 SEQ ID NO: 6).

A sliding clamp loader within the meaning of the present invention is understood to be in particular a protein that exhibits in a sequence alignment to the human (eukaryotic) amino acid sequence (SEQ ID NO: 1, 23, 33, 34) having a length of at least 100 amino acids at least a 20% sequence identity.

A sliding clamp loader 2 within the meaning of the present invention is also understood to be in particular a protein that exhibits in a sequence alignment to the human (eukaryotic)

**THIS PAGE BLANK (USPTO)**

amino acid sequence (SEQ ID NO: 6) having a length of at least 150 amino acids at least a 20% sequence identity.

Suitable clamp loaders are, for example, the Archaeobacteria-derived homologs cited in Fig. 1, (SEQ ID NO: 3, 4, 5; homologs to sliding clamp loader 1 and SEQ ID NO: 8, 9, 10, homologs to sliding clamp loader 2). Therefore, a thermostable prokaryotic in vitro complex is especially preferred, in which an RFC-homolog protein or protein complex is also present.

Thus, for example, a sliding clamp loader 1 within the meaning of the present invention is also understood to be any protein that contains the following consensus sequences and which deviates from these sequences at no more than four positions (Abb.6):

SEQ ID NO: 41

C-N-Y-X-S-[K R H D E ]-I-I-X-[G A V L I M P F W]-[G A V L I M P F W]-Q-S-R-C-X-X-F-R-F-X-P-[G A V L I M P F W].

Thus, for example, a sliding clamp loader 2 within the meaning of the present invention is also understood to be any protein that contains the following consensus sequences and which deviates from these sequences at no more than four positions (Abb. 7):

SEQ ID NO: 42

K-X-X-L-L-X-G-P-P-G-X-G-K-T-[S T N Q Y C]-X-[G A V L I M P F W]-X-X-[G A V L I M P-F W]

Additionally, a Hidden Markov Model was generated from the multiple alignment of human sliding clamp loader 1 homologs shown in Fig. 14. Thus, a sliding clamp loader 1 within the meaning of the present invention is understood to be in particular any protein that with a Hidden Markov Model so generated has a score of greater than 25 (Fig. 14).

Additionally, a Hidden Markov Model was generated from the multiple alignment of human sliding clamp loader 2 homologs shown in Fig. 15. Thus, a sliding clamp loader 2 within the meaning of the present invention is understood to be in particular any protein that with a Hidden Markov Model so generated has a score of greater than 15 (Fig. 15).

It is also preferred a sliding clamp loader is present in the form of a protein that is homologous with the Eubacterium Eschereichia coli  $\gamma$ -complex.

Also preferred in accordance with the present invention is a thermostable in vitro complex for the elongation of nucleic acids, in which apart from the sliding clamp loader, ATP may also be present, preferably also associated therewith.

#### Coupling Protein

The function of the coupling protein is to connect the elongation protein with the sliding clamp protein. A coupling protein within the meaning of the present invention is understood to

**THIS PAGE BLANK (USPTO)**

be in particular any protein having the above described function. Suitable coupling proteins are, for example, the Archaeobacteria-derived homologs to the human sequence of the coupling subunit (DPD2-HUMAN, [SEQ ID NO: 16] cited in **Fig. 1** (SEQ ID NO: 17, 18, 19, 20, 21)).

A coupling protein within the meaning of the present invention is understood to be in particular a protein that exhibits in a sequence alignment to the human (eukaryotic) amino acid sequence (SEQ ID NO: 16) having a length of at least 150 amino acids at least a 18% sequence identity.

Thus, a coupling protein within the meaning of the present invention is understood to be any protein that contains the following consensus sequence and deviates from this sequence at no more than four positions (**Fig. 5**);

SEQ ID NO: 43

[FL]-G A V L I M P F W]-X-X-[G A V L I M P F W]-X-G-X(13)-[G A V L I M P F W]-X-[YR]-[G A V L I M P F W]-X-[G A V L I M P F W]-A-G-[DN]-[G A V L I M P F W]-[G A V L I M P F W]-[DS].

A Hidden Markov Model was also generated from the multiple alignment of human coupling subunit homologs shown in **Fig. 16**. Thus, a coupling subunit within the meaning of the present invention is understood to be any protein that with a Hidden Markov Model so generated has a score of greater than 10 (**Fig. 16**).

Elongation protein exhibiting polymerase activity, or enzyme complex exhibiting polymerase activity or a subunit thereof

Some elongation proteins require the presence of a coupling protein (coupling subunit) in order to exhibit polymerase activity at all. However, it is conceivable that the elongation protein may bind directly to the sliding clamp. Other elongation proteins require the presence of a coupling protein for bonding with the slide clamp.

The elongation protein exhibits 5'-3' polymerase activity or reverse-transcriptase activity. Suitable proteins in such case are, for example, Archaeobacteria-derived homologs of the human elongation protein (SEQ ID NO: 22) cited in **Fig. 1** (SEQ ID NO: 23, 24, 25, 26).

Preferably, the elongation protein binds to the coupling protein, which in turn is coupled to the sliding clamp protein.

An elongation protein within the meaning of the present invention is understood to be in particular a protein that exhibits in a sequence alignment to the human (eukaryotic) amino acid sequence (SEQ ID NO: 22) having a length of at least 200 amino acids at least a 20% sequence identity.

**THIS PAGE BLANK (USPTO)**

Thus, an elongation protein within the meaning of the present invention is understood to be in particular any protein that contains the following consensus sequence and deviates from this sequence at no more than four positions (**Fig. 8**):

SEQ ID NO: 44

D-[G A V L I M P F W]-[G A V L I M P F W]-X-X-Y-N-X-X-X-F-D-X-P-Y-[G A V L I M P F W]-X-X-R-A.

A Hidden Markov Model was also generated from the multiple alignment of human elongation protein homologs (SEQ ID NO: 22) shown in **Fig. 17**. Thus, an elongation protein within the meaning of the present invention is understood to be any protein that with a Hidden Markov Model so generated has a score of greater than 20 (**Fig. 17**).

An elongation protein within the meaning of the present invention is understood to be in particular a protein that exhibits in a sequence alignment to the archaeabacterial amino acid sequence (SEQ ID NO: 27) having a length of at least 400 amino acids at least a 25% sequence identity. For example, suitable proteins are those derived from Archaeabacteria with the SEQ ID NO: 28, 29, 30 or 31.

Thus, an elongation protein within the meaning of the present invention is understood to be in particular any protein that contains the following consensus sequence and deviates from this sequence at no more than four positions (**Fig. 9**):

SEQ ID NO: 45

A-[G A V L I M P F W]-R-T-A[G A V L I M P F W]-A-[G A V L I M P F W]-[G A V L I M P F W]-T-E-G-[G A V L I M P F W]-V-X-A-P-[G A V L I M P F W]-E-G-I-A-X-V-[K R H D E]-I.

A Hidden Markov Model was also generated from the multiple alignment of human elongation protein homologs (SEQ ID NO: 27) shown in **Fig. 18**. Thus, an elongation protein within the meaning of the present invention is understood to be any protein that with a Hidden Markov Model so generated has a score of greater than 35 (**Fig. 18**).

The elongation protein may also be eubacterial in origin.

Thus, an elongation protein within the meaning of the present invention is also understood to be a protein that exhibits in a sequence alignment to the eubacterial amino acid sequence (SEQ ID NO: 37) having a length of at least 300 amino acids at least a 25% sequence identity.

Thus, an elongation protein within the meaning of the present invention is also understood to be in particular any protein that contains the following consensus sequence and deviates from this sequence at no more than four positions (**Fig. 10**):

**THIS PAGE BLANK (USPTO)**



SEQ ID NO: 46

[G A V L I M P F W]-P-V-G-[G A V L I M P F W]-G-R-G-S-X-[G A V L I M P F W]-G-S-[G A V L I M P F W]-V-A-X-A-[G A V L I M P F W]-X-I-T-D-[G A V L I M P F W]-D-P-[G A V L I M P F W]-X-X-X-[G A V L I M P F W]-L-F-E-R-F-L-N-P-E-R-[G A V L I M P F W]-S--M-P-D.

A Hidden Markov Model was also generated from the multiple alignment of eubacterial elongation protein homologs (SEQ ID NO: 37) shown in **Fig. 19**. Thus, an elongation protein within the meaning of the present invention is understood to be any protein that with a Hidden Markov Model so generated has a score of greater than 20 (**Fig. 19**).

Heretofore, some DNA polymerases have been used as elongation proteins without coupling proteins and without sliding clamps in Standard PCR reactions, such as, for example, DNA polymerase I derived from *Pyrococcus furiosus* (U.S. Patent No. 5,545,552) or from *Pyrococcus* species (EP Patent Application No. 0 547 359 A1). Such enzymes have the distinctive property of being thermostable and frequently of possessing 3'-5' exonuclease activity ('proof-reading activity'). Only recently was a heterodimer having polymerase activity discovered in *Pyrococcus furiosus* (Uemori, T., Sato, Y., Kato, I., Doi, H., and Ishino, y. (1997). "A novel DNA polymerase in the hyperthermophilic archaeon, *pyrococcus furiosus*: gene cloning, expression, and characterization", *Genes to Cells* 2:499-512.)

Within the scope of the present invention, it is preferable that the inventive prokaryotic in vitro complex for elongating nucleic acids consist of proteins derived from Archaea. That is, it is preferable to utilize a sliding clamp protein derived from Archaeabacteria. Further, it is preferable for the elongation protein or a protein complex exhibiting polymerase activity consisting of an elongation protein and coupling proteins, to be derived from Archaeabacteria.

Exemplars of proteins suitable for the complex according to the present invention are shown in **Figs. 1 and 2**.

Of course, it is possible to optimize the properties of said proteins by deletions or mutations, or by addition of amino acids. Such altered proteins as well form the subject matter of the present invention, to the extent that they form the in vitro complex for elongating nucleic acids in accordance with the present invention, in addition to fulfilling the functions described in greater detail above.

The complex according to the present invention is illustrated in **Fig. 3**, which shows by way of example the replication apparatus in a potential variant, wherein the sliding clamp is bonded by means of a coupling subunit to the elongation protein.

Further, it is preferable that in addition to the prokaryotic accessory in vitro complex according to the present invention, two primers are also present. As a rule, primers are oligonucleotides that bond to both complementary DNA strands of a target sequence, in which disposed in opposite orientation with their 3' ends directed towards one another, they enclose the

**THIS PAGE BLANK (USPTO)**

segment being amplified. They serve as a starting point for amplification and generally provide the polymerase with a free 3'-OH end for incorporating a nucleotide.

As the protein complex is used for amplification, elongation and sequencing, the complex according to the present invention is present preferably in a suitable buffer. Suitable buffers are those that can be utilized for PCR, sequencing, nucleic acid labeling and for other in vitro nucleic acid elongation reactions by means of polymerases. Suitable buffers are described, for example, in *Methods in Molecular Biology*, Vol. 15, Humana Press Totowa, New Jersey, 1993, edited by Bruce A. White.

As the protein complex is used for elongation, amplification, reverse transcription and/or sequencing, there is present, in addition to the complex according to the present invention, a mixture of nucleotides. Deoxynucleotides may be selected from dGTP, dATP, dTTP and dCTP, but are not limited to these. In addition, derivatives of deoxynucleotides are also feasible according to the present invention, the former being defined as those deoxynucleotides capable of being incorporated by way of a thermostable polymerase in growing DNA molecules that are being synthesized in the reaction. Such derivatives include, but are not limited to, thionucleotide, 7-deaza-2'-dGTP, 7-deaza-2'-dATP as well as deoxyinosine triphosphate, which may also be used as substitute deoxynucleotide for dATP, dGTP, dTTP or dCTP. It is also feasible to use labeled deoxynucleotides. All known forms of labeling and/or those suitable for the purpose of the present invention may be present.

Dideoxynucleotides may be selected from ddGTP, ddATP, ddTTP and ddCTP, but are not limited to these. In addition, derivatives of dideoxynucleotides are also feasible according to the present invention, the former being defined as those dideoxynucleotides capable of being incorporated by way of a thermostable polymerase in growing DNA molecules that are being synthesized in the reaction. Such derivatives include, but are not limited to, radioactive dideoxynucleotides (ddATP, ddGTP, ddTTP and ddCTP) or dideoxynucleotides (ddATP, ddGTP, ddTTP and ddCTP), that are labeled, amongst others, with, e.g. FITC, Cy5, Cy5.5, Cy7 and Texas red. Within the scope of a sequencing according to the present invention it is also feasible to use labeled deoxynucleotides together with unlabeled dideoxynucleotides.

Ribonucleotides may be selected from GTP, ATP, TTP and CTP, but are not limited to these. In addition, derivatives of ribonucleotides are also feasible according to the present invention, the former being defined as those ribonucleotides capable of being incorporated by way of a thermostable polymerase in growing DNA molecules that are being synthesized in the reaction. Such derivatives include, but are not limited to, radioactive ribonucleotides (ATP, GTP, TTP and CTP) or ribonucleotides (ATP, GTP, TTP and CTP), that are labeled, amongst others, with, e.g. FITC, Cy5, Cy5.5, Cy7 and Texas red.

When using the protein complex for amplification, elongation and sequencing, it may prove advantageous if a pyrophosphatase is present during the reaction.

Further subject matter of the present invention is a thermostable, accessory in vitro complex comprising both a sliding clamp protein and a coupling protein, wherein said proteins are defined as above.

**THIS PAGE BLANK (USPTO)**

The accessory in vitro complex according to the present invention is to a certain extent to be understood as a precursor to the above described inventive thermostable in vitro complex for template-dependent elongation of nucleic acids.

The accessory complex need only be combined with a single elongation protein in order to confer on the latter significantly enhanced processivity. Thus, it is feasible to use the accessory in vitro complex in combination with known thermostable polymerases, wherein the disadvantages of such known proteins as well, in particular with regard to processivity, are reduced.

Gene identification, gene cloning, gene expression and purification of proteins of the in vitro complex according to the present invention:

Generally, complexes consisting of recombinant proteins may be prepared according to following steps: Preparation of the nucleic acid fragment that encodes for the desired protein, ligation in an expression vector, transformation into a host, expression and purification of the protein. In accordance with the present invention, it may happen that genes, in particular Archaeobacterial derived, may contain inteins that must first be removed (Proc Natl Acad Sci USA, June 15 1992; 89(12):5577-5581, "Intervening sequences in an Archaea DNA polymerase gene", Perler, F.B., Comb, D.G., Jack, W.E., Moran, L.S., Qiang, B., Kucera, R.B., Benner, J., Slatko, B.E., Nwankwo, D.O., Hempstead, S.K., et al.)

Identification of additional proteins suitable for the complex of the present invention may be accomplished, e.g. through homology searching in databases that encompass prokaryotic genomes. Several programs are suitable for such purpose, such as for example, the program BLASTP and FASTA (Altschul, Stephen F., Thomas L. Madden, Alejandro A., Schaffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997, "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", Nucleic Acids Res. 25:3389-3402. W.R. Pearson & D.J. Lipman PNAS (1988) 85:2444-2448).

Identification may also be accomplished using DNA probes, for the purpose of screening for suitable genes in e.g. total-genomic prokaryotic databases. The experimental methods required for this are found in Maniatis et al. (Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, N.Y. (1989).

The purified nucleic acid of the genes of the inventive complex may be prepared, e.g. by isolating them from a genomic database of the relative organism or by means of synthetically manufactured DNA, in each case optimally in combination with PCR amplification with the help of primers specific to the desired gene segment. Conventional methods are described in Maniatis et al. (Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, N.Y. (1989).

The genes of the proteins of the inventive in vitro complex can be cloned in accordance with a number of techniques, and thus be made available using an expression vector for protein expression in a host organism. Current methods are described in Maniatis et al. (Molecular

**THIS PAGE BLANK (USPTO)**

Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, N.Y. (1989). Here, the genes of the complex may e.g. first be cloned in a "high-copy" vector, for example, pUC18, pBst or pBR322, and only then recloned in a prokaryotic expression vector, such as pTrc99, or alternatively directly cloned in prokaryotic expression vector. Here, vectors are understood to be nucleic acids capable of transporting a second nucleic acid molecule into or between different organisms, or between genetic backgrounds. As a rule, they have the capability of autonomic replication and/or expressing (expression vectors) the operatively bound nucleic acid molecule. "Operatively bound" means that the transported nucleic acid molecule is connected to the vector in such a way that transcription and translation thereof are controlled by expression control sequences of the vector and may be expressed in a host cell. Bacterial expression systems, their preferred use and a selection of vector systems are described, e.g. in "Gene Expression Technology", (Method. Enzymol., Vol. 185, Goeddel, Ed., Academic Press, N.Y. (1990). Vectors suitable for the present invention should enable various expression levels of the proteins by possessing some or all of the following characteristics: (1) promoters, or transcription initiation sites, either directly at the start of the protein, or as fusion protein (2) operators that may be used to switch gene expression on or off, (3) ribosomal binding sites for improved translation, and (4) transcription or translation termination sites which lead to improved stability.

Expression vectors compatible with eukaryotic cells, preferably with vertebrate cells, may also be used. Some known vectors are pSVL and pKSV-10 (Pharmacia), pBPv-1/pML2d (International Biotechnologies, Inc.) and pTDT1 (ATCC 31255). Retroviral expression vectors may also be used.

Thus, the subject matter of the present invention also includes DNA-sequences which code for thermostable in vitro complexes or accessory in vitro complexes according to the present invention, as well as corresponding vectors, preferably expression vectors.

The vectors according to the present invention contain at least the gene for the sliding clamp protein and preferably at least a gene for a coupling and/or elongation protein, as described above, respectively.

It is preferable within the scope of the present invention if in addition to DNA sequences already contained therein, the vector also contains suitable restriction interfaces and, optionally, polylinkers for inserting additional DNA sequences. It is especially preferred if the spatial arrangement of already present DNA sequence and additional insertion sites result after expression in the formation of a fusion protein.

Further, it is preferred if the vector of the present invention also contains promoters and/or operator regions, in which it is especially preferred if such promoters and/or operator regions are inducible or repressible. In this way, control of the expression in the host cells is simplified considerably, and said control may be efficiently designed.

Promoter/operator regions of this type may occur repeatedly in an expression vector, which optionally enables independent expression of multiple DNA sequences while utilizing just one expression vector.

**THIS PAGE BLANK (USPTO)**



Within the scope of the present invention, an especially preferred vector contains DNA sequences that encode for all components of an in vitro complex or accessory in vitro complex in accordance with the present invention.

The subject matter of the present invention further includes a host cell which contains one or more vectors according to the present invention, wherein proteins may be expressed in said host cell under suitable conditions. Suitable conditions include, for example, the presence of an inducer or a derepressor.

Standard protocols exist for transformation, phage infection and cell culture in Maniatis et al. (Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, N.Y.). Of the multiplicity of available E. coli strains suitable for transformation, those preferred are JM101 (ATCC No. 33876), XL1 (Stratagene), RRI (ATCC No. 31343) and BL21 (Pharmacia). Protein expression may also ensue e.g. with the help of E. coli strain INVαF' (Invitrogen).

The transformants are cultivated in accordance with suitable growing conditions of the host strain. Thus, for example, most E. coli strains are cultivated in LB medium, at 30°C to 42°C up to a logarithmic or stationary growth phase. The proteins may be purified from a transformed culture, this being effected either from a cell pellet, after centrifugation or from the culture solution as well. In the event the proteins are purified from a cell pellet, said cells are then re-suspended in a suitable buffer, then broken open by means of sonification, treatment with enzymes or freezing and thawing. For purification from a culture suspension, that is, alone or via a fusion protein, the supernatant cells are separated off using known methods, such as centrifugation.

Separation and purification of proteins of the complexes according to the present invention, either from the supernatant of the culture solution or from the cell extract are effected using known methods of separation or purification. Said methods are e.g. of the kind that involve solubility, such as salt precipitates, and solvent precipitates, methods which exploit differences of molecular weight, such as dialysis, ultra-filtration, gel filtration, and SDS polyacrylamide gel electrophoresis, methods which exploit differences in electrical charge, ion exchange chromatography, methods which exploit different hydrophobicities, such as reverse-phase HPLC (High Performance Liquid Chromatography), methods which exploit specific affinities, such as affinity chromatography, and methods which exploit differences in isoelectric point, such as isoelectric focusing. It is also contemplated that cell extracts may be made, either from the organism that carries the gene of the accessory complex, which achieves the object of the present invention, or from the recombinant host organism, for example, E. coli. Using these extractions, it would be possible under certain circumstances to avoid additional purification steps.

The methods described above may be utilized in a number of combinations for preparing the proteins of the in vitro complex.

The thermostable prokaryotic accessory in vitro complex of the present invention may be used for elongating nucleic acids, e.g. for polymerase chain reactions, DNA sequencing or labeling of nucleic acids, as well as other reactions that contain nucleic acid in vitro synthesis. A

**THIS PAGE BLANK (USPTO)**

further possibility is the application in reverse transcription, wherein either the inventive complex itself exhibits reverse transcriptase activity, or alternatively a suitable enzyme is added which exhibits reverse transcriptase activity.

Thus, the subject matter of the present invention also includes a method for template-dependent elongation of nucleic acids, in which the nucleic acids when necessary are denatured, are provided with at least one primer under conditions of hybridization, said primer being complementary to a flanking region of a desired nucleic acid sequence of the template strand, and in which primer elongation occurs with the help of a polymerase in the presence of nucleotides, and wherein a thermostable in vitro complex of the present invention is used as a polymerase.

Methods for template-dependent elongation of nucleic acids, in which elongation starts at a primer that has been hybridized to the template nucleic acid and which provides a free 3'-OH end for the elongation, are known by those skilled in the art. For amplification, a polymerase chain reaction in particular is performed. Here, one normally starts with a double-stranded DNA sequence, of which a specific target region is to be amplified. Two primers are used which are complementary to regions flanking the target sequence on a partial strand each of the DNA double strand. In order to hybridize the primers, however, the double-stranded DNA is first denatured, in particularly thermally fused. Once the primers have been hybridized, elongation occurs by means of the polymerase, the DNA is then denatured once again, and finally the newly formed DNA strands are separated from the template strands. Thereupon, both the original template strands and the new nucleic acid strands formed in step one are present as templates for a second elongation cycle. The former are in turn hybridized with primers and a second elongation occurs. This process is performed cyclically with thermal denaturing as an intermediate step.

A thermostable in vitro complex according to the present invention is also used for purposes of reverse transcription of RNA to DNA, preferred in accordance with the invention, the elongation protein of which exhibits reverse transcriptase activity. Said reverse transcriptase activity may be the sole polymerase activity exhibited by the elongation protein, but it may also be present in addition to 5'-3'-DNA-polymerase activity.

Yet another preferred method according to the present invention involves the sequencing of nucleic acids, initiated by a primer that is complementary to a region adjacent the nucleic acid being sequenced, and wherein a template-dependent elongation or, in the case of RNA sequencing, reverse transcription is performed using deoxynucleotides and dideoxynucleotides in accordance with the Sanger method. Within the scope of said preferred embodiment, the derivatives described above are also considered suitable deoxynucleotides or dideoxynucleotides, respectively. For the method of elongation of nucleic acids of the present invention, it is especially preferred if the nucleic acids so formed are labeled. For such purpose, it is feasible to use labeled primers and/or labeled deoxynucleotides and/or labeled dideoxynucleotides and/or labeled ribonucleotides or corresponding derivatives, respectively, exemplars of which were previously described above.

**THIS PAGE BLANK (USPTO)**

Still further, the subject matter of the present invention also includes a method for labeling nucleic acids by means of which single breaks are incorporated in the phosphodiester bonds of the nucleic acid chain and a nucleotide is replaced at the break sites by a labeled nucleotide with the help of a polymerase, wherein the polymerase used is a thermostable in vitro complex according to the present invention.

A method of this type, commonly called nick translation, allows for simple labeling of nucleic acids. All of the above described labeled ribonucleotides or deoxyribonucleotides or derivatives thereof are suitable for labeling, as long as the polymerase accepts them as substrate.

Labeling according to the present invention may be accomplished, for example, within the scope of or subsequent to a PCR reaction, and for which the use according to the present invention of a thermostable in vitro complex is especially favorable.

The subject matter of the present invention also includes a kit for elongation and/or amplification and/or reverse transcription and/or sequencing of nucleic acids, wherein the kit may be present in one or more containers, and contains

- a) a thermostable in vitro complex according to the present invention or
- b) a thermostable, accessory in vitro complex and, optionally, a separate elongation protein exhibiting polymerase activity and, optionally, primers, buffer substances, nucleotides, ATP, additional cofactors and/or pyrophosphate.

In particular, the subject matter of the present invention includes a kit for the elongation, amplification, reverse transcription, labeling and sequencing of nucleic acids, and which additionally contains deoxynucleotides or derivatives thereof.

Thus, for the amplification of nucleic acids, a preferred reagent kit according to the present invention, in addition to substances a) or b) which exhibit 5'-3'-polymerase activity, also contains deoxynucleotides and/or derivatives thereof. Optionally, ribonucleotides or derivatives thereof may also be used, namely when a polymerase is used that also accepts ribonucleotides as substrate.

Additional preferred subject matter of the present invention includes a kit for sequencing nucleic acids which, in addition to deoxynucleotides or derivatives thereof, also contains dideoxynucleotides or derivatives thereof for chain termination.

Further, preferred subject matter of the present invention includes in particular a kit for reverse transcription of nucleic acids in which either the complex of the present invention itself exhibits reverse transcriptase activity, or additionally, a suitable enzyme exhibiting reverse transcriptase activity is present, in which deoxynucleotides or derivatives thereof are contained in the reaction mixture.

In yet another preferred embodiment said kit contains primers and/or deoxynucleotides and/or dideoxynucleotides and/or ribonucleotides and/or respective labeled derivatives thereof.

**THIS PAGE BLANK (USPTO)**

For the sequencing of nucleic acids in particular, it is necessary to insert a label. Exemplars of suitable labels were described above and are also considered as preferred embodiments within the scope of the reagent kit according to the present invention.

It is also feasible within the scope of the present invention to utilize the reagent kit for labeling nucleic acids within the framework of a so-called nick translation. In such case, the reagent kit contains the components a) or b) and labeled nucleotides, in which buffer substances, ATP or other cofactors and/or pyrophosphate may also be present. As a rule, primers are generally not required within the scope of a nick-translation.

Further, the subject matter of the present invention also includes use of a thermostable sliding clamp protein in in-vitro methods for elongation, amplification, labeling and sequencing or reverse transcription of nucleic acids.

Said kit is especially preferred if the sliding clamp protein is present in the form of a thermostable homolog to the PCNA-protein (SEQ ID NO: 12, 13, 14, 15), Archaeoglobus Fulgidus-derived PCNA homologs, or homolog of the  $\beta$ -clamp-protein (SEQ ID NO: 35, 36). A kit is especially preferred which also contains a coupling protein.

Further, a kit is preferred which contains an additional protein, wherein the additional protein is present as a clamp loader, is a derivative of the group of prokaryotes and is thermostable.

In particular a kit is preferred that contains a buffer as described above. It is also preferable if the kit according to the present invention also contains a pyrophosphatase, ATP and/or other cofactors.

The present invention is described in connection with the figures in greater detail below:

## Figures

Frequently, sequence names are used below, by which the protein and nucleic acid sequences are identified in the Genbank and the EMBL database.

## Abb. 1

Protein sequences, aligned in pairs, and multiple alignments derived from Archaeobacteria and from corresponding human genes of the replication apparatus.

In which the annotation: <sup>1</sup> signifies % identity to the corresponding human gene, calculated from the paired alignment (see appendix) using BLASTP 2.0.4 [Feb-24-1998], and the annotation <sup>2</sup> signifies % identity to the corresponding Archaeoglobus Fulgidus-derived gene, calculated from the paired alignment (see appendix) using BLASTP 2.0.4 [Feb-24-1998], and the annotation <sup>3</sup> signifies % identity to the corresponding human gene, calculated from the paired alignment (see appendix) using FASTA 3.1t02 [March, 1998]. These procedures are described in greater detail in: Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schaffer, Jinghui

**THIS PAGE BLANK (USPTO)**



Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", *Nucleic Acids Res.* 25: 389-3402, and W.R. Pearson & D.J. Lipman, *PNAS* (1988) 85:2444-2448. The figure shows in the case of sliding clamp loader I, sliding clamp loader II, sliding clamp, coupling subunit and elongation protein I the sequence names from the databases as well as their SEQ ID NO, in which the values indicated in brackets represent in each case the percentage identity per number of amino acids. In the case of the elongation protein II the values refer to the percentage sequence identity with the *Archaeoglobus fulgidus* sequence.

**Fig. 2**

Protein sequences, aligned in pairs, and multiple alignments derived from eubacteria of the replication apparatus, in which the annotation <sup>1</sup> signifies % identity to the corresponding *E. coli*-derived gene, calculated from the paired alignment (see appendix) using BLASTP 2.0.4 [Feb-24-1998]. The procedure is described in greater detail in: Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schaffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", *Nucleic Acids Res.* 25: 389-3402.

**Fig. 3**

Outline of a potential variant of the replication apparatus, in which the sliding clamp is bonded via a coupling subunit to the elongation protein.

**Fig. 4**

**Fig. 4** shows the alignments of two conserved regions of the sliding clamp protein, and the consensus sequences derived therefrom. The following genes are shown: PCNA human (from SEQ ID NO:11), the corresponding sequence derived from *Archaeoglobus fulgidus* (from SEQ ID NO: 12), from *Methanococcus janashii* (from SEQ ID NO: 13), from *Pyrococcus horikoshii* (from SEQ ID NO: 14), and from *Methanococcus thermoautotrophicus* (from SEQ ID NO: 15).

**Fig. 5**

**Fig. 5** shows an alignment of a conserved region of the coupling subunit, and the consensus sequences derived therefrom. The following genes are shown: PfuORF2, DPD2\_HUMAN, AF1790 and MJ0702. The SEQ ID NOs may be found in Fig.1.

**Fig. 6**

**Fig. 6** shows an alignment of a conserved region of the coupling subunit, and the consensus sequences derived therefrom. The following genes are shown: AC11\_HUMAN, AF2060, MTH 0241, PHBN012 and MJ1422. The SEQ ID NOs may be found in Fig.1.

**Fig. 7**

**THIS PAGE BLANK (USPTO)**

**Fig. 7** shows an alignment of a conserved region of the sliding clamp loader 2, and the consensus sequences derived therefrom. The following genes are shown: AC15\_HUMAN, MJ0884, AF1195, MTH0240 and MTH0240. The SEQ ID NOs may be found in **Fig.1**.

**Fig. 8**

**Fig. 8** shows an alignment of a conserved region of the elongation protein 1, and the consensus sequences derived therefrom. The following genes are shown: DPOD\_HUMAN, MJO0885, MTH1208, PHBT047 and DPOL\_ARCFU. The SEQ ID NOs may be found in **Fig.1**.

**Fig. 9**

**Fig. 9** shows an alignment of a conserved region of the elongation protein 2, and the consensus sequences derived therefrom. The following genes are shown: AF1722, MJ1630, PfuORF3, MTH1536 and PHBN021. The SEQ ID NOs may be found in **Fig.1**.

**Fig. 10**

**Fig. 10** shows an alignment of a conserved region of the Eubacteria-derived elongation protein, and the consensus sequences derived therefrom. The following genes are shown: DP3A\_ECOLI: DNA Pol III, alpha subunit, Escherichia coli, BB0579: DNA Pol III, alpha subunit, Borrelia burgdorferi, DP3A\_HELPY: DNA Pol III, alpha subunit, Helicobacter pylori AA50: Aquifex aeolicus, section 50 and DP3A\_SALTY: DNA Pol III, alpha subunit, Salmonella typhimurium).

**Fig. 11**

**Fig. 10** shows an alignment of a conserved region of the Eubacteria-derived sliding clamp, and the consensus sequences derived therefrom. The following genes are shown: AAPOL3B, DP3B\_ECOLI, S.TYPHIM, DP3B\_PROMI, DP3B\_PSEPU and DP3B\_STRCO (AAPOL3B: Aquifex Aeolicus section 93: DP3B\_ECOLI: DNA Pol III, beta chain, Escherichia coli S.TYPHIM: DNA Pol III, beta chain, Salmonella typhimurium P3B\_PROMI: DNA Pol III, beta chain, Probeus mirabilis DP3B\_PSEPU: DNA Pol III, beta chain, Pseudomonas putida DP3B\_STRCO: DNA Pol III, beta chain, Strptomyces coelicolor).

**Fig. 12**

Multiple alignment of the sliding clamp protein sequence for generating the Hidden Markov Model.

**Fig. 13**

Multiple alignment of the eubacterial sliding clamp protein sequence for generating the Hidden Markov Model (AAPOL3B: Aquifex Aeolicus section 93: DP3B\_ECOLI: DNA Pol III, beta chain, Escherichia coli S.TYPHIM: DNA Pol III, beta chain, Salmonella typhimurium

**THIS PAGE BLANK (USPTO)**

P3B\_PROMI: DNA Pol III, beta chain, *Probeus mirabilis* DP3B\_PSEPU: DNA Pol III, beta chain, *Pseudomonas putida* DP3B\_STRCO: DNA Pol III, beta chain, *Strptomyces coelicolor*).

**Fig. 14**

Multiple alignment of the sliding clamp loader 1 protein sequences for generating the Hidden Markov Model.

**Fig. 15**

Multiple alignment of the sliding clamp loader 2 protein sequences for generating the Hidden Markov Model.

**Fig. 16**

Multiple alignment of the protein sequences of the coupling subunits for generating the Hidden Markov Model.

**Fig. 17**

Multiple alignment of the elongation protein 1 sequences for generating the Hidden Markov Model.

**Fig. 18**

Multiple alignment of the elongation protein 2 sequences for generating the Hidden Markov Model.

**Fig. 19**

Multiple alignment of the eubacterial elongation protein sequences for generating the Hidden Markov Model. The following genes are shown:

DP3A\_ECOLI: DNA Pol III, alpha subunit, *Escherichia coli*, BB0579: DNA Pol III, alpha subunit, *Borrelia burgdorferi*, DP3A\_HELPY: DNA Pol III, alpha subunit, *Helicobacter pylori* AA50: *Aquifex aeolicus*, section 50 and DP3A\_SALTY: DNA Pol III, alpha subunit, *Salmonella typhimurium*).

**Example**

Using known methods, the DNA is purified from the *Archaeoglobus fulgidus* organism (DSM No. 4304). Here organisms are cultured by means of the DSM (German Collection of Microorganisms). To clone the appropriate genes (sliding clamp loader I/II, sliding clamp, elongation proteins I/II, coupling subunit) in the expression vector pTrc99, primers for each gene are developed which encompass the full open reading frame, as well as start and stop codon. In

**THIS PAGE BLANK (USPTO)**

the process, restriction ends are added to the primers, which facilitate directed cloning in the expression vector. Using approximately 200 ng of total-genomic DNA, PCR reactions are driven by suitable annealing temperatures (approximately 35 cycles), and the resulting product is purified. After purification, the products are treated with restriction enzymes and purified via an agarose gel to be readied for ligation. Using restriction enzymes, the expression vector is linearized, purified and attenuated in such a way as to prepare it for ligation with the amplicants of the accessory gene from the aforementioned PCR. Ligation reaction is prepared and following incubation an aliquot is transformed in the *E. coli* strain INValphaF' (Invitrogen). Three positive colonies are picked from each gene, plasmid DNA is prepared, and the inserts are checked via DNA sequencing for completeness and accuracy. Correct clones are picked, then recovered once again for singling out on agar plates (ampicillin). Colonies are then picked and overnight cultures prepared. An aliquot (500 µl) of the overnight culture is placed in a 1 to 5l culture of LB (Ampicilin: 80- mg/l). The cultures grow until they reach an OD<sub>660</sub> of 0.8 at 37°C, at which point IPTG (125 mg/l) is added for induction. The cultures continue to grow for an additional 11 hours. The cultures are then centrifuged and the pellets are absorbed in a buffer (Buffer A: 50 mM Tris-HCL pH 7.9, 50 mM dextrose, 1 mM EDTA). Following centrifugation, the cells are re-absorbed, but now Buffer A also contains Lysozyme (4 mg/ml). Following incubation (15 min), an equal volume of Buffer B is added (B: 10 mM Tris-HCL pH 7.9, 50 mM KCl, 1 mM EDTA, 1 mM PMSF, 0.5 % Tween 20, 0.5% Nonidet P40) and the lyse is added for incubation at 75°C for one hour. Following centrifugation, the supernatant is removed and the overexpressed proteins are precipitated out using (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>. Following centrifugation, the pellets are collected and the proteins are re-suspended with Buffer A. The re-suspended proteins are dialyzed against storage buffer (50 mM Tris-HCL pH 7.9, 50 mM KCl, 0.1 mM EDTA, 1 mM DTT, 0.5 mM PMSF, 50% glycerol) and subsequently stored at -70°C.

To test the activity of the proteins, reactions are composed as follows: Aliquots of the proteins are combined in different configurations and molarities, sliding clamp loader I/II with sliding clamp, coupling subunit and elongation protein I, or sliding clamp loader I/II with sliding clamp, with and without coupling subunit and elongation protein II, or sliding clamp and elongation protein I or II, and finally just elongation protein I or II; the buffer is the aforementioned storage buffer. DNA polymerase activity is measured by incorporating (methyl-<sup>3</sup>H) TTP in trichloric acid-insoluble material (Ishino, Y., Iwasaki, H., Fukui, H., Mineno, J., Kato, I., & Schinigawa, H. (1992) *Biochimie* 74:131-136). To determine processivity the aforementioned protein mixtures are used in primer-elongation experiments. A M13 single-stranded template is immersed in 10 mM Tris-HCl (pH 9.4) and heated (92°C) together with a universal primer (labeled 5'-FITC), then cooled (room temperature). Attenuation series of the template-primer mix thus generated are concentrated in a reaction consisting of nucleotides (approximately 200 µM to 1 mM), reaction buffer (final concentration: 50 mM KCl, 10 mM Tris-HCl ) pH 8.3), 1.5-5 mM MgCl<sub>2</sub>, ATP (0 mM-200 mM) and protein stabilizing agents, and incubated for 10 minutes at 37°C, 52°C, 62°C, 68°C, 74°C, 78°C. One aliquot is loaded for analysis on an automatic sequencer (e.g. Alf, Pharmacie Biotech). The aforementioned protein mixtures are also used to measure fidelity and exonuclease activity, in which the method described in Kohler et al. is used (*Proc. Natl. Acad. Sci. USA* 88:7958-7962 (1991) or Chase et al. (*J. Biol. Chem.*, 249: 4545-4552 (1972)). Said protein mixtures are also used in PCR (*Methods in Molecular Biology*, Vol. 15, Humana Press, Totowa, New Jersey, 1993, edited by Bruce A. White).

**THIS PAGE BLANK (USPTO)**



## Claims

1. Thermostable in vitro complex for template-dependent elongation of nucleic acids, comprising a thermostable sliding clamp protein which is connected with an elongation protein that exhibits polymerase activity.
2. Thermostable complex according to claim 1, characterized in that the sliding clamp protein and the elongation protein are connected via a coupling protein.
3. Thermostable complex according to claim 1 or 2, characterized in that the sliding clamp protein and the elongation protein are derived from archaeobacteria.
4. Thermostable complex according to claim 1 or 2 or 3, characterized in that said sliding clamp protein has a ring-like structure the encircles entirely or in part the template-nucleic acid strands.
5. Thermostable complex according to any of claims 1-4, characterized in that the sliding clamp protein has one or both of the following consensus sequences:  
[G A V L I M P F W]-D-X-X-X-[G A V L I M P F W]-X-X-[G A V L I M P F W]-X-[G A V L I M P F W]-X-[G A V L I M P F W]-X-X-X-F-X-X-Y-X-X-D and/or  
[G A V L I M P F W]-X(3)-L-A-P-[K R H D E]-[G A V L I M P F W]-E.
6. Thermostable complex according to any of claims 1-5, characterized in that the sliding clamp protein exhibits in a sequence alignment to the human (eukaryotes) PCNA amino acid sequence (SEQ ID NO: 11) having a length of at least 100 amino acids at least a 20% sequence identity, and/or the sliding clamp protein exhibits in a sequence alignment to the bacterial  $\beta$ -clamp sequence derived from *E. coli* (eubacteria) (SEQ ID NO: 35) having a length of at least 100 amino acids at least a 20% sequence identity, and/or the sliding clamp protein exhibits in a sequence alignment to the amino acid sequence of the PCNA homologs derived from *Archaeoglobus Fulgidus* (Archaea) (SEQ ID NO: 12) having a length of at least 100 amino acids at least a 20% sequence identity.
7. Thermostable complex according to any of the preceding claims, characterized in that the sliding clamp protein with a Hidden Markov Model generated from an alignment in Fig. 12 yields a score of at least 20, and/or in that the sliding clamp protein with a Hidden Markov Model generated from an alignment in Fig. 13 yields a score of at least 25.
8. Thermostable complex according to any of the preceding claims, characterized in that the sliding clamp protein is selected from *Archaeoglobus Fulgidus*-derived AF0335, from *Methanococcus jannaschii*-derived MJ0247, from *Pyrococcus horikoshii*-derived PHLA 008, from *Methanobacterium thermoautotrophicus*-derived MTH1312 and from *Aquifex aeolicus*-derived AE000761\_7
9. Thermostable complex according to any of the preceding claims, characterized in that the elongation protein exhibit 5'-3'-polymerase activity or reverse transcriptase activity.

**THIS PAGE BLANK (USPTO)**

10. Thermostable complex according to claim 9, characterized in that the elongation protein contains at least one of the following consensus sequences and deviates from said sequence at no more than four positions:

SEQ ID NO: 44

D-[G A V L I M P F W]-[G A V L I M P F W]-X-X-Y-N-X-X-X-F-D-X-P-Y-[G A V L I M P F W]-X-X-R-A

SEQ ID NO: 45

A-[G A V L I M P F W]-R-T-A-[G A V L I M P F W]-A-[G A V L I M P F W]-[G A V L I M P F W]-T-E-G-[G A V L I M P F W]-V-X-A-P-[G A V L I M P F W]-E-G-I-A-X-V-[K R H D E]-I.

SEQ ID NO: 46

[G A V L I M P F W]-P-V-G-[G A V L I M P F W]-G-R-G-S-X-[G A V L I M P F W]-G-S-[G A V L I M P F W]-V-A-X-A-[G A V L I M P F W]-X-I-T-D-[G A V L I M P F W]-D-P-[G A V L I M P F W]-X-X-X-[G A V L I M P F W]-L-F-E-R-F-L-N-P-E-R-[G A V L I M P F W]-S--M-P-D.

11. Thermostable complex according to claim 9, characterized in that the elongation protein exhibits in a sequence alignment to the human (eukaryotic) amino acid sequence (SEQ ID NO: 22) having a length of at least 200 amino acids at least a 20% sequence identity and/or in a sequence alignment to the archaeobacterial amino acid sequence (SEQ ID NO: 27) having a length of at least 400 amino acids at least a 25% sequence identity and/or in a sequence alignment to the eubacterial amino acid sequence (SEQ ID NO: 37) having a length of at least 300 amino acids at least a 25% sequence identity.

12. Thermostable complex according to claim 9, characterized in that the elongation protein with a Hidden Markov Model generated from an alignment in Fig. 17 yields a score of at least 20, and/or in that the sliding clamp protein with a Hidden Markov Model generated from an alignment in Fig. 18 yields a score of at least 35 and/or in that the sliding clamp protein with a Hidden Markov Model generated from an alignment in Fig. 19 yields a score of at least 20.

13. Thermostable complex according to any of the preceding claims, characterized in that the elongation protein is selected from Archaeoglobus Fulgidus-derived AF0497 or AF1722, from Methanococcus jannaschii-derived MJ0885 or MJ163, from Pyrococcus horikoshii-derived PHBT047 or PHBN021, from Methanobacterium thermoautotrophicus-derived MTH1208 or MTH1536 and from Pyrococcus furiosus-derived PFUORF3.

14. Thermostable complex according to any of the preceding claims, characterized in that the coupling protein contains the following consensus sequence and deviates from said sequence at no more than four positions:

(SEQ ID NO:43)

[FL]-G A V L I M P F W]-X-X-[G A V L I M P F W]-X-G-X(13)-[G A V L I M P F W]-X-[YR]-[G A V L I M P F W]-X-[G A V L I M P F W]-A-G-[DN]-[G A V L I M P F W]-[G A V L I M P F W]-[DS].

**THIS PAGE BLANK (USPTO)**

15. Thermostable complex according to any of the preceding claims, characterized in that the coupling protein exhibits in a sequence alignment to the human (eukaryotic) amino acid sequence (SEQ ID NO: 16) having a length of at least 150 amino acids at least a 18% sequence identity.
16. Thermostable complex according to any of the preceding claims, characterized in that the coupling protein with a Hidden Markow Model generated from an alignment from Fig. 16 yields a score of at least 10.
17. Thermostable complex according to any of the preceding claims, characterized in that the coupling protein is selected from Archaeoglobus Fulgidus-derived AF1790, from Methanococcus jannaschii-derived MJ0702, from Pyrococcus horikoshii-derived PHBN023, from Methanobacterium thermoautotrophicus-derived MTH1405 and from Pyrococcus furiosus-derived PFUORF2.
18. Thermostable complex according to any of the preceding claims, characterized in that the complex is associated with a protein that functions as a sliding clamp loader.
19. Thermostable complex according to any of the preceding claims, characterized in that the presence of said complex is associated with ATP or another cofactor.
20. Thermostable accessory in vitro complex, characterized in that it contains a sliding clamp protein and a coupling proteins as defined, respectively, in any of the preceding claims.
21. Recombinant DNA-sequence, characterized in that it encodes for a thermostable in vitro complex according to any of claims 1 to 19 or for a thermostable accessory complex according to claim 20.
22. Vector, characterized in that it contains a recombinant DNA sequence that encodes for sliding clamp protein and a coupling protein and/or an elongation protein.
23. Vector according to claim 22, characterized in that it also contains suitable restriction interfaces for inserting additional DNA sequences, so arranged that a fusion protein is produced from the sliding clamp protein and the expression product of the additional DNA sequences.
24. Vector according to claim 22 or 23, characterized in that it contains suitable promoter and/or operator regions which control expression of the DNA sequence(s)
25. Vector according to claim 24, characterized in that it contains multiple promoter and/or operator regions for separate expression of multiple DNA sequences.
26. Vector according to any of claims 22 to 25, characterized in that it contains repressible and inducible promoter and/or operator regions.

**THIS PAGE BLANK (USPTO)**

27. Vector according to any of claims 22 to 26, characterized in that it contains a DNA sequence in accordance with claim 21.
28. Host cell, characterized in that it is transformed with one or more vectors according to any of claims 22 to 27.
29. Method for producing a thermostable in vitro complex according to any of claims 1 to 19, or a thermostable, accessory in vitro complex according to claim 20, characterized in that a suitable recombinant DNA sequence according to claim 21 or one or more corresponding vectors according to any of claims 22 to 27 are imparted in a host cell, causing the proteins to be expressed, said expressed proteins are then isolated from culture medium or after opening of the cell, and optionally coupled with additional components of the complex.
30. Use of a thermostable, accessory in vitro complex according to claim 20 for producing a thermostable in vitro complex according to any of claims 1 to 19, characterized in that the accessory complex is coupled with an elongation protein exhibiting polymerase activity.
31. Method of template-dependent elongation of nucleic acids, in which the nucleic acids are, if necessary, denatured, provided with at least one primer under hybridizing conditions, in which said primer is complementary to a flanking region of a desired nucleic acid sequence of the template strand, and in which primer elongation occurs with the help of a polymerase in the presence of nucleotides, characterized in that the polymerase used is a thermostable in vitro complex according to any of claims 1 to 19.
32. Method according to claim 31, characterized in that two primers flanking the desired nucleic acid and deoxynucleotides and/or derivatives thereof and/or ribonucleotides or derivatives thereof are used for amplifying DNA sequences.
33. Method according to claim 32, characterized in that a polymerase chain reaction is carried out.
34. Method according to claim 31, characterized in that for reverse transcription of RNA to DNA a thermostable in vitro complex according to any of claims 1 to 19 is used, the elongation protein of which exhibits reverse-transcriptase activity.
35. Method according to any of claims 31 to 34, characterized in that sequencing of nucleic acids initiated by a primer which is complementary to a region adjacent to the nucleic acid being sequenced is carried out by way of a template-dependent elongation or reverse transcription using deoxynucleotides and dideoxynucleotides or derivatives thereof in accordance with the Sanger method.
36. Method according to any of claims 31 to 35, characterized in that labels are inserted during elongation of the nucleic acids.

**THIS PAGE BLANK (USPTO)**



37. Method according to claim 36, characterized by the use of labeled primers and/or labeled deoxynucleotides and/or derivatives thereof and/or labeled dideoxynucleotides and/or derivatives thereof and/or labeled ribonucleotides and/or derivatives thereof.
38. Method for labeling nucleic acids by producing single breaks in phosphodiester bonds of the nucleic acid chain and with the help of a polymerase, replacing the nucleotide at the break sites with a labeled nucleotide, characterized in that the polymerase used is a thermostable in vitro complex according to any of claims 1 to 19.
39. Reagent kit for elongation and/or amplification and/or reverse transcription and/or sequencing and/or labeling of nucleic acids, containing in a single or in multiple containers
- a) a thermostable in vitro complex according to any of claims 1 to 19 or
  - b) a thermostable accessory in vitro complex according to claim 20 and optionally, a separate elongation protein exhibiting polymerase activity,
- and, optionally, primers, buffer substances, nucleotides, ATP, one or more cofactors and/or pyrophosphate.
40. Kit according to claim 39, characterized in that for the purpose of nucleic acid amplification it contains in addition to substances a) or b), which exhibit 5'-3'-polymerase activity, deoxynucleotides and/or derivatives thereof.
41. Kit according to claim 39, characterized in that for the purpose of reverse transcription it contains substances a) or b), which exhibit 5'-3'-reverse transcriptase activity, as well as deoxynucleotides and/or derivatives thereof.
42. Kit according to any of claims 39 to 41, characterized in that for the purpose of sequencing it contains in addition to deoxynucleotides or ribonucleotides and/or derivatives thereof, dideoxynucleotides and/or derivatives thereof.
43. Kit according to any of claims 39 to 42, characterized in that it contains primers and/or deoxynucleotides and/or dideoxynucleotides in labeled form.

**THIS PAGE BLANK (USPTO)**



①9 BUNDESREPUBLIK  
DEUTSCHLAND



DEUTSCHES  
PATENT- UND  
MARKENAMT

⑫ **Offenlegungsschrift**  
⑩ **DE 198 40 771 A 1**

⑤① Int. Cl. 7:  
**C 12 N 9/12**  
C 07 K 19/00  
C 12 P 19/34  
C 12 N 15/62  
C 12 N 15/63  
C 12 Q 1/68

⑦① Aktenzeichen: 198 40 771.8  
⑦② Anmeldetag: 7. 9. 1998  
⑦③ Offenlegungstag: 10. 2. 2000

DE 198 40 771 A 1

⑥⑥ Innere Priorität:  
198 35 653. 6 06. 08. 1998

⑦① Anmelder:  
Lion Bioscience AG, 69120 Heidelberg, DE

⑦④ Vertreter:  
BOEHMERT & BOEHMERT, 28209 Bremen

⑦② Erfinder:  
Voss, Hartmut, Dr., 69221 Dossenheim, DE;  
Moeckel, Gerd, Dr., 68723 Oftersheim, DE; Kober,  
Ingo, Dr., 69251 Gaiberg, DE; Kilger, Christian, Dr.,  
69121 Heidelberg, DE

Die folgenden Angaben sind den vom Anmelder eingereichten Unterlagen entnommen

⑤④ Thermostabiler in vitro-Komplex mit Polymeraseaktivität

⑤⑦ Ein erfindungsgemäßer thermostabiler in vitro-Komplex zur Template-abhängigen Elongation von Nukleinsäuren umfaßt ein thermostabiles Gleitklammerprotein, welches mit einer thermostabilen Polymeraseaktivität aufweisenden Elongationsprotein verbunden ist. Ein erfindungsgemäßer thermostabiler akzessorischer in vitro-Komplex enthält ein Gleitklammerprotein und ein Kopplungsprotein und kann mit geeigneten Polymerasen kombiniert werden. Die erfindungsgemäßen in vitro-Komplexe können in Verfahren zur Template-abhängigen Elongation von Nukleinsäuren, insbesondere z. B. der PCR-Reaktion, eingesetzt werden. Entsprechende Reagenzienkits werden ebenfalls beschrieben.

DE 198 40 771 A 1

Die Erfindung betrifft einen thermostabilen in vitro-Komplex zur Template-abhängigen Elongation von Nukleinsäuren, einen thermostabilen prokaryontischen akzessorischen in vitro Komplex sowie dafür codierende DNA-Sequenzen und Vektoren. Die Erfindung betrifft weiterhin die Verwendung der erfindungsgemäßen Komplexe in Verfahren zur Template-abhängigen Elongation von Nukleinsäuren, wie PCR-Reaktionen oder der DNA Sequenzierung, bei denen in vitro Template-abhängige DNA Strangsynthese erfolgt. Schließlich betrifft die Erfindung noch Reagenzienkits zur Durchführung der erfindungsgemäßen Verfahren.

DNA Polymerasen gehören zu einer Gruppe von Enzymen, die einzelsträngige DNA als Template für die Synthese eines komplementären DNA Stranges verwenden. Diese Enzyme spielen eine bedeutende Rolle im Nukleinsäurestoffwechsel, einschließlich der Prozesse DNA Replikation, Reparatur und Rekombination. DNA Polymerasen wurden in allen zellulären Organismen identifiziert, von bakteriellen bis zu menschlichen Zellen, in vielen Viren sowie in Bakteriophagen (Kornberg, A. & Baker, T. A. (1991) DNA Replikation WH Freeman, New York, NY). Man faßt in der Regel die Archaeobakterien und die Eubakterien zusammen zu der Gruppe der Prokaryonten, der Organismen ohne echtem Zellkern, und stellt ihnen die Eukaryonten, die Organismen mit echtem Zellkern, gegenüber. Gemeinsam sind vielen Polymerasen aus den verschiedensten Organismen oftmals Ähnlichkeiten in der Aminosäuresequenz sowie Ähnlichkeiten in der Struktur (Wang, J., Sattar, A.K.M.A.; Wang, C.C., Karam, J.D., Konigsberg, W.H. & Steitz, T.A. (1997) Crystal Structure of pol a family replication DNA polymerase from bacteriophage RB69. Cell 89, 1087-1099). Organismen wie der Mensch besitzen eine Vielzahl von DNA abhängigen Polymerasen, von denen jedoch nicht alle für die DNA Replikation zuständig sind, sondern einige auch DNA Reparatur durchführen. Replikative DNA Polymerasen bestehen meist aus Protein komplexen mit mehreren Untereinheiten, welche die Chromosomen der zellulären Organismen und Viren replizieren. Eine generelle Eigenschaft dieser replizierenden Polymerasen ist im allgemeinen eine hohe Prozessivität, das heißt, deren Fähigkeit, Tausende von Nukleotide zu polymerisieren ohne vom DNA Template abzus dissoziieren (Kornberg, A. & Baker, T. A. (1991) DNA Replikation. WH Freeman, New York, NY).

Bis vor kurzem waren die einzigen gut verstandenen, hochprozessiven Replikations-Mechanismen in Zellen solche, die von zellulären Replikasen (einem Protein-Komplex mit Polymerase-Aktivität) verwendet wurden, sowie der Replikationsapparat des Bakteriophagen T4 oder T7. Der dem Mechanismus zugrunde liegende Apparat beinhaltet ein Protein mit ringähnlicher Struktur, eine "Gleitklammer", welche die DNA umschließt und die katalytische Polymeraseeinheit – das "Elongationsprotein" – an die DNA bindet (Stukenberg, P.T., Studwell-Vaughan, P.S. & O'Donnell, M. (1991) Mechanism of the  $\beta$ -clamp of DNA polymerase III holoenzyme. J. Biol. Chem. 266, 11328-11334; Kuriyan, J. & O'Donnell, M. (1993) Sliding clamps of DNA polymerases. J. Mol. Biol. 234, 915-925). Die Gleitklammer ist häufig über ein oder mehrere weitere Proteine, den "Kopplungsproteinen", an das Elongationsprotein gebunden. Die dreidimensionale Struktur von verschiedenen Gleitklammerproteinen wurde bereits bestimmt:

- die des eukaryontischen proliferating cell nuclear antigen (PCNA) (Krishna, T.S.R., Kong, X.-P., Gary, S., Burgers, P. M. & Kuriyan, J. (1994) Crystal structure of the eukaryotic DNA polymerase processivity factor PCNA. Cell 79, 1233-1243; Gulbis, J. M., Kelman, Z., Hurwitz, J., O'Donnell, M. & Kuriyan, J. (1996) Structure of the C-terminal region of p21WAF1/CIP1 complexed with human PCNA. Cell 87, 297-306),
- die der  $\beta$  Untereinheit der Polymerase III des Eubakteriums *Escherichia coli* (Kong, X.-P., Onrust, R., O'Donnell, M. & Kuriyan, J. (1992) Three dimensional structure of the  $\beta$  subunit of *Escherichia coli* DNA polymerase III holoenzyme; a sliding DNA clamp. Cell 69, 425-437)
- und die des Bakteriophagen T4 Gen45 Proteins (Kelman, Zvi, Hurwitz, J. O'Donnell, Mike (1998) Structure, 6, 121-125).

Die Gesamtstruktur dieser Gleitklammern ist sehr ähnlich; die Bilder der Proteingesamtstruktur von PCNA, der  $\beta$ -Untereinheit und der gp45 Ringe sind übereinandergelegt deckungsgleich (Kelman, Z. & O'Donnell, M. (1995) Structural and functional similarities of prokaryotic and eukaryotic sliding clamps. Nucleic Acids Res. 23, 3613-3620). Jeder Ring hat vergleichbare Dimensionen und eine zentrale Öffnung, die groß genug ist, um Duplex-DNA, also einen DNA Doppelstrang, bestehend aus den zwei komplementären DNA Strängen, zu umschließen.

Die Gleitklammer kann sich in vivo nicht selbst um die DNA herum positionieren; sie muß durch einen Klammerlader um die DNA fixiert werden. Dieser Klammerlader ist ein Protein komplex, der in Prokaryonten und Eukaryonten aus einer Vielzahl von Untereinheiten besteht und der beim Eubakterium *Escherichia coli*  $\gamma$ -Komplex beziehungsweise beim Menschen Replikationsfaktor C (RF-C) genannt wird (Kelman, Z. & O'Donnell, M. (1994) DNA replication – enzymology and mechanisms. Curr. Opin. Genet. Dev. 4, 185-195). Der Gleitklammerlader erkennt das 3'-Ende des Einzelstrang-Duplexes (Primer-Template) und positioniert die Gleitklammer in Anwesenheit von ATP um die DNA. Die Gleitklammer, welche die DNA dann umschließt, wechselwirkt mit Polymerasen, und gewährleistet so eine schnelle und prozessive DNA Synthese.

Im Falle des Bakteriophagen T7 wird das gleiche Ziel, eine prozessive DNA Synthese, mittels eines strukturell anderen Proteinkomplexes erreicht. Der Phage exprimiert eine eigene katalytische Polymerase, die T7 Polymerase, das Produkt des gene 5, welche mit einem Protein aus dem Wirt *Escherichia coli*, dem Thioredoxin eine Bindung eingeht und als Replikase eine hochprozessive DNA Replikation ermöglicht (Proc Natl Acad Sci USA 1992 Oct 15; 89(20): 9774-9778 Genetic analysis of the interaction between bacteriophage T7 DNA polymerase and *Escherichia coli* thioredoxin, Hima-wan JS, Richardson CC). Auch hierbei kommt es zur Klammerbildung, jedoch weist diese Klammer nicht die gleiche Struktur auf, wie z. B. im Falle des eukaryontischen PCNA.

Oft ist es nötig – wie zum Beispiel im Falle der humanen Polymerase  $\delta$  –, daß Proteine (Kopplungsproteine) die Verbindung zwischen dem katalytisch aktiven Teil der Polymerase und dem Prozessivitätsfaktor (Gleitklammer) schaffen. Beim Menschen ist dies die kleine Untereinheit der  $\delta$ -Polymerase (Zhang, S.-J., Zeng, X.-R., Zhang, P., Toomey, N.L., Chuang, R.Y., Chang, L.-S., and Lee, M.Y.W.T. (1994). A conserved region in the amino terminus of DNA polymerase  $\delta$

is involved in proliferating cell nuclear antigen binding. J. Biol. Chem. 270, 7988-7992). Im Falle der T7 Polymerase jedoch bindet der Prozessivitätsfaktor die katalytische Einheit der Polymerase direkt.

DNA Polymerasen werden unter anderem durch zwei Eigenschaften charakterisiert, ihre Elongationsrate, das heißt die Anzahl der Nukleotide die sie pro Sekunde in einen wachsenden DNA Strang inkorporieren können und ihre Dissoziationskonstante. Wenn die Polymerase nach jedem Inkorporationsschritt eines der Nukleotide in die wachsende Kette wieder vom Strang abdissoziiert, (d. h. ein Elongationsschritt erfolgt pro Bindungsereignis), dann hat die Prozessivität den Wert 1 und die Polymerase ist nicht prozessiv. Wenn die Polymerase für wiederholte Nukleinsäureinkorporationen mit dem Strang verbunden bleibt, dann wird der Replikationsmodus als prozessiv bezeichnet und kann einen Wert von mehreren Tausend erreichen (siehe hierzu auch: Methods in Enzymology Volume 262, DNA Replication, Edited by J. L. Campbell, Academic press 1995, pp. 270-280).

Für die meisten in vitro Anwendungen, wie PCR oder Sequenzierungsprozesse ist Prozessivität eine wünschenswerte Eigenschaft, die allerdings die bislang in diesen Reaktionen eingesetzten thermostabilen Enzyme nur in geringem Maße besitzen, wohingegen die temperatursensitive, mit Thioredoxin assoziierte T7 Polymerase eine Prozessivität von einigen tausend Nukleotiden hat. Im Vergleich – die thermostabile DNA Polymerase aus *Thermus thermophilus* oder *aquaticus* haben nur eine Prozessivität von etwa 50 Nukleotiden (Biochim Biophys Acta 1995 Nov 7; 1264(2): 243-248 Inactivation of the 5'-3' exonuclease of *Thermus aquaticus* DNA polymerase. Merckens LS, Bryan SK, Moses RE).

Die U.S. Patente 4,683,195, 4,800,195 und 4,683,202 beschreiben die Anwendung solcher thermostabilen DNA Polymerasen in der Polymerase Kettenreaktion (PCR). In der PCR wird unter Verwendung von Primern, Template, Nukleotiden, einer DNA Polymerase eines entsprechenden Puffers und geeigneten Reaktionsbedingungen DNA neu synthetisiert. Hierbei wird die doppelsträngige Zielsequenz zumeist thermisch aufgeschmolzen, zwei Oligonukleotide werden anhybridisiert und die Komplementärsequenz mittels der Inkorporation von Nukleotiden durch die Polymerase am Template synthetisiert. Das Extensionsprodukt jedes Primers dient als Template für den nächsten Zyklus. In dieser PCR kommt es bevorzugt zur Verwendung einer thermostabilen Polymerase, welche das zyklische, thermische Aufschmelzen der DNA Stränge übersteht. So wird häufig Taq DNA Polymerase verwendet (U.S. Patent 4,965,188). Die Prozessivität der Taq DNA Polymerase ist jedoch wie oben ausgeführt relativ gering im Vergleich zur T7 Polymerase.

DNA Polymerasen finden auch in der DNA Sequenzbestimmung Anwendung (Sanger et al., Proc. Natl. Acad. Sci., USA 74: 5463-5467 (1977)). Häufig wird bei der Sequenzierung nach Sanger eine T7 DNA Polymerase verwendet (Tabor, S. und Richardson, C.C. Proc. Natl. Acad. Sci., USA 86: 4076-4080 (1989)). Später wurde das Cycle-Sequencing-Verfahren entwickelt (Murray, V. (1989) Nucleic Acids Res. 17, 8889), welches kein einzelsträngiges Template erfordert und die Initiierung der Sequenzreaktion mit verhältnismäßig geringen Mengen an Template erlaubt. Die hierbei zur Verwendung kommenden Polymerasen können z. B. die oben erwähnte Taq Polymerase sein (U.S. Patent 5,075,216) oder die Polymerase von *Thermotoga neapolitana* (WO 96/10640) oder andere thermostabile Polymerasen. Neuere Verfahren koppeln die exponentielle Amplifikation und die Sequenzierung eines DNA Fragmentes in einem Schritt, so daß es möglich ist, genomische DNA direkt zu sequenzieren. Eines der Verfahren, das sogenannte DEXAS-Verfahren (Nucleic Acids Res 1997 May 15; 25(10): 2032-2034 Direct DNA sequence determination from total genomic DNA. Kilger C, Pääbo S, Biol Chem 1997 Feb; 378(2): 99-105 Direct exponential amplification and sequencing (DEXAS) of genomic DNA. Kilger C, Pääbo S und DE 196 53 439.9 sowie DE 196 53 494.1) verwendet eine Polymerase mit verminderter Diskriminierungsfähigkeit gegenüber Dideoxynukleotiden (ddNTPs) im Vergleich zu Deoxynukleotiden (dNTPs) sowie einen Reaktionspuffer, zwei Primer, die preferentiell nicht equimolar vorliegen, und die oben genannten Nukleotide, um dann in mehreren Zyklen eine komplette, sequenzspezifische DNA Leiter eines Fragmentes zu erhalten welches von den Primern umspannt ist. Eine Weiterentwicklung dieses Verfahrens besteht in der Verwendung eines Polymerasegemisches, wobei eine der beiden Polymerase zwischen ddNTPs und dNTPs diskriminiert, während die zweite eine verminderte Diskriminierungsfähigkeit aufweist (Nucleic Acids Res 1997 May 15; 25(10): 2032-2034 Direct DNA sequence determination from total genomic DNA. Kilger C, Pääbo S).

DNA Polymerasen finden auch Anwendung in der reversen Transkription von RNA in DNA. Hierbei dient RNA als Template und die Polymerase synthetisiert einen komplementären DNA Strang. Zur Anwendung kommt hier z. B. die thermostabile DNA Polymerase aus dem Organismus *Thermus thermophilus* (Tth) (U.S. Patent 5,322,770).

Es kann zudem erwünscht sein, daß die Polymerase eine 'proof-reading' Aktivität besitzt, also eine 3'-5' Exonukleaseaktivität aufweist. Diese Eigenschaft ist insbesondere dann wünschenswert, wenn das zu synthetisierende Produkt mit einer niedrigen Fehlerrate bei der Nukleotidinkorporation hergestellt werden soll.

Die oben genannten Enzyme, die üblicherweise in PCR-Reaktionen eingesetzt werden, gehören größtenteils nicht zu den eigentlichen Replikationsenzymen, sondern es sind zumeist Enzyme, von denen man annimmt, daß sie an der DNA Reparatur beteiligt sind, weshalb deren Prozessivität relativ gering ist.

Somit war es Aufgabe der vorliegenden Erfindung, mehrere der vorgenannten Eigenschaften von Polymerasen, insbesondere hohe Prozessivität und Thermostabilität für die Verwendung in in vitro Reaktionen zu vereinen.

Diese Aufgabe wurde erfindungsgemäß gelöst durch die Bereitstellung eines thermostabilen in vitro-Komplexes zur Template-abhängigen Elongation von Nukleinsäuren, umfassend ein thermostabiles Gleitklammerprotein, welches mit einem thermostabilen Polymeraseaktivität-aufweisenden Elongationsprotein verbunden ist. Dieser Komplex kann in in vitro-Reaktionen, wie z. B. in PCR-Reaktionen, eingesetzt werden und weist dabei eine hohe Prozessivität auf. Vorteil ist zudem, wenn der Komplex eine geringe Fehlerquote bei der Nukleotidinkorporation aufweist, also eine erhöhte Fidelity hat. Dieser Komplex kann somit bei der Elongation, der Amplifikation und der Sequenzierung von Nukleinsäuren eingesetzt werden. Dieser Komplex ist vorzugsweise in Standard-PCR-Reaktionen einsetzbar.

Um die Verwendbarkeit eines solchen Komplexes in Standard-PCR-Reaktionen zu ermöglichen, muß eine einfache Handhabung gewährleistet sein. So wurde zwar der Replikationsapparat des Simian virus 40 bereits in vitro zusammengesetzt (Wage, S. & Stillman, B. (1994) Nature, Vol. 369, 207-221), jedoch für Standard-PCR-Reaktionen ist dieser Replikationsapparat nicht geeignet, da er unter anderem nicht thermostabil ist. Es ist somit Gegenstand der Erfindung, einen thermostabilen in vitro-Komplex mit Polymeraseaktivität bereitzustellen, der in Standard-PCR-Reaktionen einsetzbar ist und der eine hohe Prozessivität aufweist.

Insbesondere ist Gegenstand der Erfindung ein thermostabiler prokaryontischer in vitro Komplex zur Elongation von Nukleinsäuren, der ein thermostabiles Gleitklammerprotein, welches die komplementären Nukleinsäurestränge ganz oder teilweise umschließt, und ein thermostabiles, Polymeraseaktivität aufweisendes Protein umfaßt, wobei dieses Protein oder dieser Proteinkomplex mit dem Gleitklammerprotein gekoppelt ist.

Im Rahmen der vorliegenden Erfindung umfaßt der Begriff Polymeraseaktivität-aufweisendes Elongationsprotein auch Polymeraseaktivität-aufweisende Proteinkomplexe oder Untereinheiten solcher Komplexe, welche die Polymeraseaktivität tragen.

Thermostabil im Sinne der vorliegenden Erfindung bedeutet, daß der akzessorische Komplex mit hoher Prozessivität Nukleotide in wachsende Nukleinsäurestränge inkorporiert sowohl bei niedrigen als auch bei hohen Temperaturen, die in der PCR oder einer anderen Reaktion auftreten, wie z. B. der DNA Sequenzierung.

Die PCR besteht z. B. in der Regel aus den Schritten der Denaturierung (70°C bis 98°C), dem Annealing (40°C bis 78°C) und der DNA Strangsynthese (60°C bis 76°C). Somit muß dieser Komplex mindestens zwischen ca. 60°C und ca. 70°C, insbesondere zwischen 60°C und 76°C und besonders bevorzugterweise funktionsfähig sein zwischen 40°C und 98°C. Es dürfen während der gesamten Reaktion keine irreversiblen Denaturierungserscheinungen des Komplexes oder einzelner Komponenten auftreten, welche die Elongationsreaktion unterbinden oder inhibieren.

Die Kopplung zwischen Gleitklammerprotein und Polymeraseaktivität aufweisendem Elongationsprotein kann durch kovalente, aber auch durch nicht-kovalente Bindung erfolgen. In einer bevorzugten Ausführungsform sind Gleitklammerprotein und Elongationsprotein über ein Kopplungsprotein verbunden.

In einer weiteren bevorzugten Ausführungsform der vorliegenden Erfindung stammen im erfindungsgemäßen thermostabilen in vitro Komplex die assoziierten Proteine aus Archaeobakterien. Es ist allerdings im Rahmen der vorliegenden Erfindung ebenfalls möglich, daß die assoziierten Proteine aus Eubakterien stammen. Gegenstand der vorliegenden Erfindung ist ebenfalls ein erfindungsgemäßer thermostabiler in vitro Komplex, bei dem die assoziierten Proteine zum Teil aus Archaeobakterien und zum Teil aus Eubakterien stammen.

Im Sinne der vorliegenden Erfindung umfaßt der Begriff prokaryontisches Protein sowohl Proteine aus Archaeobakterien und Proteine aus Eubakterien. Es ist bekannt, daß der Replikationsapparat in Archaea dem des eukaryontischen Replikationsapparates ähnlich ist, obwohl die Genomorganisation in Eukaryonten und Archaea gänzlich verschieden ist und die zelluläre Struktur der Eubakterien dem der Archaea ähnelt. (Edgell, D.R. and Doolittle, W.F. (1997). Archaea and the origin(s) of DNA replication proteins. Cell 89, 995-998).

Desweiteren ist eine bevorzugte Ausführungsform der vorliegenden Erfindung ein thermostabiler prokaryontischer in vitro Komplex, bei dem ein Polymeraseaktivität-aufweisender Proteinkomplex vorliegt, der aus einem Kopplungsprotein und einem oder mehreren Polymeraseaktivität-aufweisenden Elongationsproteinen besteht. Bevorzugt ist desweiteren ein erfindungsgemäßer thermostabiler prokaryontischer in vitro Komplex, bei dem das Gleitklammerprotein eine ringförmige Struktur aufweist, welches die komplementären Nukleinsäurestränge ganz oder teilweise umschließt.

#### Die Gleitklammer

Die folgenden Ausführungen sollen dazu dienen, die Funktion und die möglichen Erscheinungsformen des Gleitklammerproteins besser zu verstehen.

Das Gleitklammerprotein erfüllt die Funktion, die Polymeraseaktivität an die DNA zu binden. Entweder umschließt das Gleitklammerprotein selbst die DNA ganz oder teilweise oder durch Assoziation an das Polymeraseaktivität aufweisende Protein beziehungsweise an den Polymeraseaktivität-aufweisenden Proteinkomplex oder seine Untereinheit wird eine Klammer gebildet. In jedem Fall wird durch diese Klammerbildung die Prozessivität signifikant gesteigert, mindestens um das eineinhalbfache.

Das heißt, der erfindungsgemäße in vitro Komplex besitzt eine mindestens eineinhalbfache Prozessivität im Vergleich zum Elongationsprotein alleine, beziehungsweise im Vergleich zu einem Polymeraseaktivität aufweisenden Protein komplex ohne Gleitklammer oder einer Untereinheit davon.

Als Gleitklammer können beispielsweise Homologe des "Proliferating Cell Nuclear Antigen"-Proteinkomplex aus dem humanen Genom, oder Homologe des ebenfalls ringförmigen "β-clamp"-Proteinkomplex aus E. coli dienen, welche aus thermostabilen Organismen stammen und somit thermostabil sind, oder aus nicht thermostabilen Organismen stammen, und nachträglich durch Veränderung der Aminosäuresequenz thermostabil gemacht wurden (Eijssink VG, van der Zee JR, van den Burg B, Vriend G, Venema G, FEBS Lett 1991 Apr 22; 282(1): 13-16, Improving the thermostability of the neutral protease of Bacillus stearothermophilus by replacing a buried asparagine by leucine, Bertus Van den Burg, Gert Vriend, Oene R. Veltman, Gerard Venema, and Vincent G. H. Eijssink Engineering an enzyme to resist boiling PNAS 1998, 95: 2056-2060). Dabei kann die Gleitklammer aus mehreren Komponenten aufgebaut sein. Die im menschlichen Genom identifizierte Gleitklammer besteht aus drei PCNA-Protein Komponenten (SEQ ID NO: 11) (Homotrimeres), die im E. coli Genom identifizierte Gleitklammer besteht aus zwei Komponenten (SEQ ID NO: 35) (Homodimeres).

Als Gleitklammer im Sinne der vorliegenden Erfindung ist hierbei insbesondere jedes Protein zu verstehen, das die funktionelle Eigenschaft der Polymeraseprozessivitätssteigerung besitzt oder der Senkung der Fehlerrate dient. Dazu kann die Gleitklammer eine ringförmige dreidimensionale Struktur aufweisen oder durch Kopplung an ein anderes Protein ringförmige dreidimensionale Struktur bilden, durch die sie in der Lage ist, ein- und doppelsträngige DNA ganz oder teilweise zu umschließen.

Als Gleitklammer im Sinne der vorliegenden Erfindung ist insbesondere ein solches Protein zu verstehen, das

1. zu der menschlichen (Eukaryonten) PCNA Aminosäuresequenz (SEQ ID NO: 11) auf einer Länge von mindestens 100 Aminosäuren bei einem Sequenzalignment eine mindestens 20%ige Sequenzidentität aufweist oder das
2. zu der bakteriellen β-clamp Sequenz aus E. coli (Eubakteria) (SEQ ID NO: 35) auf einer Länge von mindestens 100 Aminosäuren bei einem Sequenzalignment eine mindestens 20%ige Sequenzidentität aufweist oder das
3. zu der Aminosäuresequenz des PCNA Homologen aus Archaeoglobus Fulgidus (Archaea) (SEQ ID NO: 12) auf

einer Länge von mindestens 100 Aminosäuren bei einem Sequenzalignment eine mindestens 20%ige Sequenzidentität aufweist.

Die erfindungsgemäße Gleitklammer kann eines oder mehrere der vorgenannten Merkmale aufweisen.

Die in Abb. 1 aufgeführten Sequenzidentitäten wurden mit dem BLAST Algorithmus nach Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J., J. Mol. Biol. 215, 403-410 (1990) ermittelt und werden weiter unten erläutert.

Als Gleitklammer im Sinne der vorliegenden Erfindung sind außerdem Proteine zu verstehen, die eine oder beide der folgenden Konsensussequenzen beinhalten und an nicht mehr als vier Positionen von einer dieser Sequenzen abweichen (Abb. 4):

#### Region 1

(SEQ ID NO: 39)

[G A V L I M P F W]-D-X-X-X-[G A V L I M P F W]-X-X-[G A V L I M P F W]-X-[G A V L I M P F W]-X-X-X-X-F-X-X-Y-X-X-D und/oder

#### Region 2

(SEQ ID NO: 40)

[G A V L I M P F W]-X(3)-L-A-P-[K R H D E]-[G A V L I M P F W]-E.

Die Aminosäuren werden hierbei gemäß der Standard IUPAC Einbuchstaben Nomenklatur benannt und gemäß dem Prosile Pattern Beschreibungsstandard aufgeführt. Dabei sind die folgenden Aminosäuregruppen häufig zusammengefaßt:

G, A, V, L, I, M, P, F oder W (Aminosäuren mit nicht polaren Seitenketten)

S, T, N, Q, Y, oder C (Aminosäure mit ungeladenen polaren Seitenketten)

K, R, H, D oder E (Aminosäure mit geladenen und polaren Seitenketten)

Außerdem bedeutet X in den Sequenzprotokollen jede beliebige Aminosäure oder Insertion oder Deletion.

Aus dem in Abb. 12 dargestellten multiplen Alignment von E. coli  $\beta$ -clamp Homologen wurde zudem ein Hidden Markov Modell generiert. Als Gleitklammer im Sinne der vorliegenden Erfindung ist somit insbesondere jedes Protein zu verstehen, daß mit dem so erzeugten Hidden Markov Modell einen Score mehr als 20 aufweist (Abb. 12). Die Hidden Markov Modelle und die entsprechenden Scores wurden mit dem hmmfs Programm (Version 1.8.4, July 1997) aus dem HMMER Paket berechnet (HMMER Protein and DNA Hidden Markov Models (Version 1.8) von Sean Eddy, Dept. of Genetics, Washington University School of Medicine, St. Louis, USA).

Aus dem in Abb. 13 dargestellten multiplen Alignment von E. coli  $\beta$ -clamp Homologen wurde ein Hidden Markov Modell generiert. Als Gleitklammer im Sinne der vorliegenden Erfindung ist somit insbesondere jedes Protein zu verstehen, das mit dem so erzeugten Hidden Markov Modell einen Score von mehr als 25 aufweist (Abb. 13).

Die Gleitklammer kann aus mehreren Komponenten aufgebaut sein, die durch eine charakteristische Bindung fest aneinander gebunden sind, so daß ein stabiler ringförmiger Molekülkomplex gebildet wird, der nicht ohne Weiteres von der DNA dissoziieren kann. Dadurch wird eine feste aber nicht kovalente Bindung an die DNA ermöglicht, die freie Verschiebbarkeit auf derselben aber nicht behindert. Die prozessivitätssteigernden Gleitklammerproteine haben zudem charakteristische lokale Moleküleigenschaften im Bereich der Wechselwirkungsregion zur DNA, welche die freie Verschiebbarkeit erleichtern und die durch in diese Region eingelagerte Wassermoleküle unterstützt werden kann.

Eine bevorzugte Ausführungsform der vorliegenden Erfindung ist weiterhin insbesondere ein thermostabiler prokaryontischer in vitro Komplex, wobei das Gleitklammerprotein eines der folgenden ist: AF0335 aus Archaeoglobus Fulgidus, MJ0247 aus Methanococcus Jannaschii, PHLA008 aus Pyrococcus Horikoshii, MTH1312 aus Methanobacterium Thermoautotrophicus sowie AE000761\_7 aus Aquifex Aeolicus.

Insbesondere sind thermostabile prokaryontische in vitro Komplexe vom Gegenstand dieser Anmeldung umfaßt, wobei das Gleitklammerprotein eine Aminosäuresequenz gemäß SEQ ID No. 11, 12, 13, 14, 15 und 36 (Aquifex Aeolicus) aufweist.

#### Der Gleitklammerlader

Desweiteren ist als bevorzugte Ausführungsform zu verstehen, wenn zu diesem erfindungsgemäßen Komplex zusätzlich ein Gleitklammerlader gehört, welcher die Komponenten der Gleitklammer um den ununterbrochenen DNA-Strang herum zusammenfügt, bzw. diese wieder entfernt, wenn die Reaktion beendet wird. Dieser Gleitklammerlader ist bevorzugt mit dem erfindungsgemäßen in vitro-Komplex assoziiert.

Im Menschen besteht der Gleitklammerlader aus fünf Untereinheiten, 4 kleinen (Gleitklammerlader 1) und einer großen Untereinheit (Gleitklammerlader 2). Als Gleitklammerlader dienen erfindungsgemäß beispielsweise ein oder mehrere prokaryontische Homologe des im Menschen identifizierten "Replication Factor C"-Proteinkomplexes (Gleitklammerlader 1: SEQ ID NO: 1, 32, 33, 34 und Gleitklammerlader 2 SEQ ID NO: 6).

Als Gleitklammerlader 1 im Sinne der vorliegenden Erfindung ist insbesondere ein solches Protein zu verstehen, das zu der menschlichen (Eukaryonten) Aminosäuresequenz (SEQ ID NO: 1, 32, 33, 34) auf einer Länge von mindestens 100 Aminosäuren bei einem Sequenzalignment eine mindestens 20%ige Sequenzidentität aufweist.

Als Gleitklammerlader 2 im Sinne der vorliegenden Erfindung ist insbesondere auch ein solches Protein zu verstehen,

das zu der menschlichen (Eukaryonten) Aminosäuresequenz (SEQ ID NO: 6) auf einer Länge von mindestens 150 Aminosäuren bei einem Sequenzalignment eine mindestens 20%ige Sequenzidentität aufweist.

Als Klammerlader sind beispielsweise die in **Abb. 1** aufgeführten Homologe aus Archaeobakterien geeignet (SEQ ID NO: 3, 4, 5; Homologe zu Gleitklammerlader 1 und SEQ ID NO: 8, 9, 10, Homologe zu Gleitklammerlader 2). Daher ist insbesondere bevorzugt ein thermostabiler prokaryotischer in vitro Komplex, wobei zusätzlich ein RFC-homologes Protein bzw. Proteinkomplex anwesend ist.

Als Gleitklammerlader 1 im Sinne der vorliegenden Erfindung ist daher beispielsweise jedes Protein zu verstehen, das die folgenden Konsensussequenzen beinhaltet und an nicht mehr als vier Positionen von dieser Sequenz abweicht (**Abb. 6**):

SEQ ID NO: 41

C-N-Y-X-S-[K R H D E]-I-I-X-[G A V L I M P F W]-[G A V L I M P F W]-Q-S-R-C-X-X-F-R-F-X-P-[G A V L I M P F W].

Als Gleitklammerlader 2 im Sinne der vorliegenden Erfindung ist daher beispielsweise auch jedes Protein zu verstehen, das die folgenden Konsensussequenzen beinhaltet und an nicht mehr als vier Positionen von dieser Sequenz abweicht (**Abb. 7**):

SEQ ID NO: 42

K-X-X-L-L-X-G-P-P-G-X-G-K-T-[S T N Q Y C]-X-[G A V L I M P F W]-X-X-[G A V L I M P F W].

Aus dem in **Abb. 14** dargestellten multiplen Alignment von menschlichen Gleitklammerlader 1 Homologen wurde zudem ein Hidden Markov Modell generiert. Als Gleitklammerlader 1 im Sinne der vorliegenden Erfindung ist somit jedes Protein zu verstehen, das mit dem so erzeugten Hidden Markov Modell einen Score mehr als 25 aufweist (**Abb. 14**).

Aus dem in **Abb. 15** dargestellten multiplen Alignment von menschlichen Gleitklammerlader 2 Homologen wurde zudem ein Hidden Markov Modell generiert. Als Gleitklammerlader 2 im Sinne der vorliegenden Erfindung ist somit jedes Protein zu verstehen, das mit dem so erzeugten Hidden Markov Modell einen Score mehr als 15 aufweist (**Abb. 15**).

Bevorzugt ist ebenfalls, wenn ein dem Eubakterium *Escherichia coli*  $\gamma$ -Komplex homologes Protein als Gleitklammerlader anwesend ist.

Außerdem bevorzugt ist ein erfindungsgemäßer thermostabiler in vitro Komplex zur Elongation von Nukleinsäuren, bei dem neben dem Gleitklammerlader zusätzlich ATP vorliegen kann, vorzugsweise ebenfalls assoziiert.

#### Das Kopplungsprotein

Das Kopplungsprotein hat die Funktion, das Elongationsprotein und das Gleitklammerprotein zu verbinden. Als Kopplungsprotein im Sinne der vorliegenden Erfindung ist insbesondere jedes Protein zu verstehen, welches die oben beschriebene Funktion besitzt. Dabei sind als Kopplungsproteine beispielsweise die in **Abb. 1** aufgeführten Homologe zu der humanen Sequenz der koppelnden Untereinheit (DPD2\_HUMAN, [SEQ ID NO: 16]) aus Archaeobakterien geeignet (SEQ ID NO: 17, 18, 19, 20, 21).

Als Kopplungsprotein im Sinne der vorliegenden Erfindung ist insbesondere ein solches Protein zu verstehen, das zu der menschlichen (Eukaryonten) Aminosäuresequenz (SEQ ID NO: 16) auf einer Länge von mindestens 150 Aminosäuren bei einem Sequenzalignment eine mindestens 18%ige Sequenzidentität aufweist.

Als Kopplungsprotein im Sinne der vorliegenden Erfindung ist daher jedes Protein zu verstehen, das die folgende Konsensussequenz beinhaltet und an nicht mehr als vier Positionen von dieser Sequenz abweicht (**Abb. 5**):

SEQ ID NO: 43

[FL]-[G A V L I M P F W]-X-X-[G A V L I M P F W]-X-G-X(13)-[G A V L I M P F W]-X-[YR]-[G A V L I M P F W]-X-[G A V L I M P F W]-A-G-[DN]-[G A V L I M P F W]-[G A V L I M P F W]-[DS].

Aus dem in **Abb. 16** dargestellten multiplen Alignment von Homologen zur menschlichen koppelnden Untereinheit wurde zudem ein Hidden Markov Modell generiert. Als koppelnde Untereinheit im Sinne der vorliegenden Erfindung ist somit jedes Protein zu verstehen, das mit dem so erzeugten Hidden Markov Modell einen Score mehr als 10 aufweist (**Abb. 16**).

Polymeraseaktivität-aufweisendes Elongationsprotein bzw. Polymeraseaktivität aufweisender Enzymkomplex bzw. Untereinheit davon

Einige Elongationsproteine benötigen die Anwesenheit eines Kopplungsproteins (koppelnde Untereinheit), um überhaupt Polymeraseaktivität aufzuweisen. Es ist aber auch vorstellbar, daß das Elongationsprotein direkt an die Gleitklammer bindet. Andere Elongationsproteine benötigen die Anwesenheit eines Kopplungsproteins für die Bindung an die Gleitklammer.

Das Elongationsprotein weist eine 5'-3'-Polymeraseaktivität oder eine Reverse-Transkriptase-Aktivität auf. Dabei sind als Elongationsproteine beispielsweise die in **Abb. 1** aufgeführten zum menschlichen Elongationsprotein (SEQ ID NO: 22) Homologe aus Archaeobakterien geeignet (SEQ ID NO: 23, 24, 25, 26).

Bevorzugterweise bindet das Elongationsprotein an das Kopplungsprotein, welches wiederum an das Gleitklammer-



protein gekoppelt ist.

Als Elongationsprotein im Sinne der vorliegenden Erfindung ist insbesondere ein solches Protein zu verstehen, das zu der menschlichen (Eukaryonten) Aminosäuresequenz (SEQ ID NO: 22) auf einer Länge von mindestens 200 Aminosäuren bei einem Sequenzalignment eine mindestens 20%ige Sequenzidentität aufweist.

Als Elongationsprotein im Sinne der vorliegenden Erfindung ist daher insbesondere jedes Protein zu verstehen, das die folgende Konsensussequenz beinhaltet und an nicht mehr als vier Positionen von dieser Sequenz abweicht (Abb. 8):

SEQ ID NO: 44

D-[G A V L I M P F W]-[G A V L I M P F W]-X-X-Y-N-X-X-X-F-D-X-P-Y-[G A V L I M P F W]-X-X-R-A.

Aus dem in Abb. 17 dargestellten multiplen Alignment von Homologen zum menschlichen Elongationsprotein (SEQ ID NO: 22) wurde zudem ein Hidden Markov Modell generiert. Als Elongationsprotein im Sinne der vorliegenden Erfindung ist somit insbesondere jedes Protein zu verstehen, das mit dem so erzeugten Hidden Markov Modell einen Score mehr 20 aufweist (Abb. 17).

Als Elongationsprotein im Sinne der vorliegenden Erfindung ist insbesondere auch ein solches Protein zu verstehen, das zu der archaeabakteriellen Aminosäuresequenz (SEQ ID NO: 27) auf einer Länge von mindestens 400 Aminosäuren bei einem Sequenzalignment eine mindestens 25%ige Sequenzidentität aufweist. Beispielsweise sind geeignet die aus Archaeabakterien stammenden Proteine mit der SEQ ID NO: 28, 29, 30 oder 31.

Als Elongationsprotein im Sinne der vorliegenden Erfindung ist daher insbesondere auch jedes Protein zu verstehen, das die folgende Konsensussequenz beinhaltet und an nicht mehr als vier Positionen von dieser Sequenz abweicht (Abb. 9):

SEQ ID NO: 45

A-[G A V L I M P F W]-R-T-A-[G A V L I M P F W]-A-[G A V L I M P F W]-[G A V L I M P F W]-T-E-G-[G A V L I M P F W]-V-X-A-P-[G A V L I M P F W]-E-G-I-A-X-V-[K R H D E]-I.

Aus dem in Abb. 18 dargestellten multiplen Alignment von Homologen zum archabakteriellen Elongationsprotein (SEQ ID NO: 27) wurde zudem ein Hidden Markov Modell generiert. Als Elongationsprotein im Sinne der vorliegenden Erfindung ist somit insbesondere jedes Protein zu verstehen, das mit dem so erzeugten Hidden Markov Modell einen Score mehr als 35 aufweist (Abb. 18).

Das Elongationsprotein kann auch eubakteriellen Ursprungs sein.

Als Elongationsprotein im Sinne der vorliegenden Erfindung ist somit ebenfalls ein solches Protein zu verstehen, das zu der eubakteriellen Aminosäuresequenz (SEQ ID NO: 37) auf einer Länge von mindestens 300 Aminosäuren bei einem Sequenzalignment eine mindestens 25%ige Sequenzidentität aufweist.

Als Elongationsprotein im Sinne der vorliegenden Erfindung ist daher auch insbesondere jedes Protein zu verstehen, das die folgende Konsensussequenz beinhaltet und an nicht mehr als acht Positionen von dieser Sequenz abweicht (Abb. 10):

SEQ ID NO: 46

[G A V L I M P F W]-P-V-G-[G A V L I M P F W]-G-R-G-S-X-[G A V L I M P F W]-G-S-[G A V L I M P F W]-V-A-X-A-[G A V L I M P F W]-X-I-T-D-[G A V L I M P F W]-D-P-[G A V L I M P F W]-X-X-X-[G A V L I M P F W]-L-F-E-R-F-L-N-P-E-R-[G A V L I M P F W]-S-M-P-D.

Aus dem in Abb. 19 dargestellten multiplen Alignment von Homologen zum eubakteriellen Elongationsprotein (SEQ ID NO: 37) wurde zudem ein Hidden Markov Modell generiert. Als Elongationsprotein im Sinne der vorliegenden Erfindung ist somit jedes Protein zu verstehen, das mit dem so erzeugten Hidden Markov Modell einen Score mehr 20 aufweist (Abb. 19).

Bisher wurden einige DNA Polymerasen als Elongationsproteine ohne Kopplungsprotein und ohne Gleitklammer für Standard PCR Reaktionen eingesetzt, so z. B. DNA Polymerase I aus *Pyrococcus furiosus* (United States Patent No. 5,545,552) oder *Pyrococcus species* (European Patent Application No: 0 547 359 A1). Diese Enzyme zeichnen sich durch die Eigenschaft aus, thermostabil zu sein und häufig eine 3'-5' exonuklease Aktivität ('proof-reading' Aktivität) zu besitzen. Erst vor kurzem wurde ein Heterodimer mit Polymeraseaktivität in *Pyrococcus furiosus* entdeckt (Uemori, T., Sato, Y., Kato, I., Doi, H., and Ishino, Y. (1997). A novel DNA polymerase in the hyperthermophilic archaeon, *pyrococcus furiosus*: gene cloning, expression, and characterization. *Genes to Cells* 2, 499-512.)

Bevorzugt ist im Rahmen der vorliegenden Erfindung, daß der erfindungsgemäße prokaryontische in vitro Komplex zur Elongation von Nukleinsäuren aus Proteinen besteht, die aus Archaea stammen. Das heißt, bevorzugt ist die Verwendung eines Gleitklammerproteins aus Archaeabakterien. Desweiteren ist bevorzugt, daß das Elongationsprotein beziehungsweise ein Polymeraseaktivität-aufweisender Proteinkomplex, welcher aus Elongationsprotein und Kopplungsprotein besteht, aus Archaeabakterien stammt.

Beispiele solcher für den erfindungsgemäßen Komplex geeigneter Proteine sind in Abb. 1 und 2 dargestellt.

Natürlich ist es möglich, durch Deletionen oder Mutationen oder durch das Anfügen von Aminosäuren die Eigenschaften dieser Proteine zu optimieren. Diese veränderten Proteine sind ebenfalls Gegenstand der vorliegenden Erfindung, solange sie den erfindungsgemäßen in vitro Komplex zur Elongation von Nukleinsäuren bilden und die oben näher spezifizierten Funktionen erfüllen.

Zur Veranschaulichung des erfindungsgemäßen Komplexes dient Abb. 3, die beispielhaft den Replikationsapparat in

einer möglichen Variante darstellt, wobei die Gleitkammer über eine koppelnde Untereinheit an das Elongationsprotein bindet.

Desweiteren ist bevorzugt, daß zusätzlich zu dem erfindungsgemäßen prokaryotischen akzessorischen in vitro Komplex zwei Primer anwesend sind. Primer sind in der Regel Oligonukleotide, welche an die beiden komplementären DNA Stränge der Zielsequenz binden, wobei sie in gegensätzlicher Orientierung, ihre 3'-Enden zueinander gerichtet, den zu amplifizierenden Abschnitt einschließen. Sie dienen als Startpunkt der Amplifikation und stellen in der Regel ein freies 3'-OH Ende für die Polymerase zur Inkorporation eines Nukleotides zur Verfügung.

Während der Verwendung des Proteinkomplexes zur Amplifikation, Elongation und Sequenzierung liegt der erfindungsgemäße Komplex bevorzugterweise in einem geeigneten Puffer vor. Geeignete Puffer sind solche, die für PCR, Sequenzierung, Nukleinsäuremarkierung und anderen in vitro Nukleinsäure Elongationsreaktionen mittels Polymerase Anwendung finden. Geeignete Puffer werden beispielweise beschrieben in *Methods in Molecular Biology* Vol. 15 Humana Press Totowa, New Jersey, 1993, edited by Bruce A. White.

Während der Verwendung des Proteinkomplexes zur Elongation, Amplifikation, Reversen Transkription oder/und Sequenzierung liegt zusätzlich zu dem erfindungsgemäßen Komplex ein Gemisch aus Nukleotiden vor. Deoxynukleotide können aus dGTP, dATP, dTTP und dCTP ausgewählt werden, sind jedoch nicht auf diese beschränkt. Zusätzlich können auch Derivate von Deoxynukleotiden gemäß der Erfindung verwendet werden, welche als solche Deoxynukleotide definiert sind, die in der Lage sind, durch eine thermostabile DNA Polymerase in wachsende DNA Moleküle inkorporiert zu werden, die in der Reaktion synthetisiert werden. Solche Derivate schließen Thionukleotide, 7-deaza-2'-dGTP, 7-deaza-2'-dATP sowie Deoxyinosine Triphosphat, das auch als Ersatz Deoxynukleotid für dATP, dGTP, dTTP oder dCTP verwendet werden kann, ein, sind aber nicht auf diese beschränkt. Ebenso können markierte Deoxynukleotide verwendet werden. Alle bekannten oder/und zu dem erfindungsgemäßen Zweck geeigneten Markierungen können dabei vorliegen.

Dideoxynukleotide können aus ddGTP, ddATP, ddTTP und ddCTP ausgewählt werden, sind jedoch nicht auf diese beschränkt. Zusätzlich können auch Derivate von Dideoxynukleotiden gemäß der Erfindung verwendet werden, die als solche Dideoxynukleotide definiert sind, die in der Lage sind, durch eine thermostabile DNA Polymerase in wachsende DNA Moleküle inkorporiert zu werden, die in der Reaktion synthetisiert werden. Solche Derivate können radioaktive Dideoxynukleotide (ddATP, ddGTP, ddTTP und ddCTP) oder Dideoxynukleotide (ddATP, ddGTP, ddTTP und ddCTP), welche mit z. B. FITC, Cy5, Cy5.5, Cy7 und Texas-Rot oder anderen markiert sind, einschließen, sind aber nicht auf diese beschränkt. Im Rahmen einer erfindungsgemäßen Sequenzierung können auch markierte Deoxynukleotide zusammen mit unmarkierten Dideoxynukleotiden eingesetzt werden.

Ribonukleotide können aus GTP, ATP, TTP und CTP ausgewählt werden, sind jedoch nicht auf diese beschränkt. Zusätzlich können auch Derivate von Ribonukleotiden gemäß der Erfindung verwendet werden, die als solche Ribonukleotide definiert sind, die in der Lage sind, durch eine thermostabile DNA Polymerase in wachsende DNA Moleküle inkorporiert zu werden, die in der Reaktion synthetisiert werden. Solche Derivate können radioaktive Ribonukleotide (ATP, GTP, TTP und CTP) oder Ribonukleotide (ATP, GTP, TTP und CTP), welche mit z. B. FITC, Cy5, Cy5.5, Cy7 und Texas-Rot oder anderen markiert sind, einschließen, sind aber nicht auf diese beschränkt.

Während der Verwendung des Proteinkomplexes zur Amplifikation, Elongation und Sequenzierung kann es sich als vorteilhaft erweisen wenn bei der Reaktion eine Pyrophosphatase anwesend ist.

Ein weiterer Gegenstand der vorliegenden Erfindung ist ein thermostabiler, akzessorischer in vitro-Komplex, welcher ein Gleitkammerprotein und ein Kopplungsprotein umfaßt, wobei beide Proteine wie oben definiert sind.

Der erfindungsgemäße akzessorische in vitro-Komplex ist gewissermaßen als Vorstufe zu dem oben beschriebenen erfindungsgemäßen thermostabilen in vitro-Komplex zur Template-abhängigen Elongation von Nukleinsäuren zu verstehen.

Der akzessorische Komplex muß lediglich mit einem Elongationsprotein kombiniert werden, um diese in eine deutlich erhöhte Prozessivität zu verleihen. Der akzessorische in vitro-Komplex kann daher auch zusammen mit bekannten thermostabilen Polymerasen eingesetzt werden, wobei ebenfalls Nachteile dieser bekannten Proteine hinsichtlich insbesondere der Prozessivität verringert werden können.

Identifizierung der Gene, Klonierung der Gene, Expression dieser und Reinigung der Proteine der erfindungsgemäßen in vitro-Komplexe:

In der Regel können die Komplexe, bestehend aus rekombinanten Proteinen, mit den folgenden Schritten bereitgestellt werden: Bereitstellung des Nukleinsäurefragmentes welches für das gewünschte Protein kodiert, Ligation in einen Expressionsvektor, Transformation in einen Wirt, Expression und Aufreinigung des Proteins. In Übereinstimmung mit der vorliegenden Erfindung kann es vorkommen, daß Gene, insbesondere aus den Archaeobakterien, Inteine (*Proc Natl Acad Sci USA* 1992 Jun 15; 89(12): 5577-5581, Intervening sequences in an Archaea DNA polymerase gene, Perler FB, Comb DG, Jack WE, Moran LS, Qiang B, Kucera RB, Benner J, Slatko BE, Nwankwo DO, Hempstead SK, et al) enthalten können, welche zunächst entfernt werden können.

Die Identifizierung weiterer für den erfindungsgemäßen Komplex geeigneten Proteine kann z. B. erfolgen durch Homologiesuchen in Datenbanken, welche Genome aus Prokaryonten umfassen. Einige Programme sind hierfür geeignet, beispielsweise das Programm BLASTP und FASTA (Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schaffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", *Nucleic Acids Res.* 25: 3389-3402. W.R. Pearson & D.J. Lipman *PNAS* (1988) 85: 2444-2448).

Die Identifizierung kann auch dadurch geschehen, daß DNA Sonden verwendet werden, um in z. B. gesamtgenomischen Banken aus Prokaryonten nach den entsprechenden Genen zu screenen. Die hierfür nötigen experimentellen Verfahren finden sich in Maniatis et al. (*Molecular Cloning: A laboratory Manual*, Cold Spring Harbor Laboratory Press, N.Y.(1989)).

Die Bereitstellung der gereinigten Nukleinsäure der Gene der erfindungsgemäßen Komplexe kann z. B. über deren Isolierung aus einer genomischen Bank des relevanten Organismus geschehen oder durch synthetische DNA-Herstellung, jeweils gewünschtenfalls kombiniert mit einer Amplifikation mittels PCR, unter Zuhilfenahme von Primern, wel-

che für den gewünschten Genabschnitt spezifisch sind. Übliche Verfahren sind in Maniatis et al. (Molecular Cloning: A laboratory Manual, Cold Spring Harbor Laboratory Press, N.Y. (1989)) beschrieben.

Die Gene der Proteine der erfindungsgemäßen in vitro-Komplexe können in einer Vielzahl von Verfahren kloniert werden und somit mittels eines Expressionsvektors zur Proteinexpression in einem Wirtsorganismus zur Verfügung gestellt werden. Gängige Verfahren sind in Maniatis et al. (Molecular Cloning: A laboratory Manual, Cold Spring Harbor Laboratory Press, N.Y.(1989)) beschrieben. Die Gene der Komplexe können hierbei z. B. zunächst in einen 'high-copy' Vektor, z. B. pUC18, pBst oder pBR322 kloniert werden und dann erst in einen prokaryontischen Expressionsvektor, z. B. pTrc99 umklontiert werden, oder aber direkt in einen prokaryontischen Expressionsvektor kloniert werden. Unter Vektoren sind hierbei Nukleinsäuren zu verstehen, die in der Lage sind, ein anderes Nukleinsäuremolekül in oder zwischen verschiedenen Organismen, bzw. genetischen Hintergründen zu transportieren. Sie haben in der Regel die Fähigkeit der autonomen Replikation und/oder der Expression (Expressionsvektoren) des operativ verbundenen Nukleinsäuremoleküls. "Operativ verbunden" bedeutet, daß das transportierte Nukleinsäuremolekül so mit dem Vektor verbunden ist, daß es unter der Transkriptions- und Translationskontrolle von Expressionskontrollsequenzen des Vektors steht und in einer Wirtszelle exprimiert werden kann. Bakterielle Expressionssysteme, die bevorzugte Verwendung derer, sowie eine Auswahl an Vektorsystemen ist z. B. in 'Gene Expression Technology', (Meth. Enzymol., Vol 185, Goeddel, Ed., Academic Press, N.Y. (1990)) beschrieben. Geeignete Vektoren für die vorliegende Erfindung sollten unterschiedlich starke Expression der Proteine dadurch ermöglichen, daß sie einige oder alle der folgenden Eigenschaften besitzen: (1) Promotoren, oder Transkriptionsinitiationsstellen, entweder unmittelbar neben dem Start des Proteins oder als Fusionsprotein, (2) Operatoren die verwendet werden können Genexpression an oder aus zu schalten, (3) ribosomale Bindungsstellen für eine verbesserte Translation, und (4) Terminationsstellen für die Transkription oder Translation, die zu verbesserter Stabilität führen.

Expressionsvektoren, die mit eukaryontischen Zellen, bevorzugterweise mit Vertebratenzellen kompatibel sind, können auch Verwendung finden. Einige bekannte Vektoren sind pSVL und pKSV-10 (Pharmacia), pBPV-1/pML2d (International Biotechnologies, Inc.), und pTDT1 (ATCC 31255). Die Verwendung eines retroviralen Expressionsvektors ist auch möglich.

Weitere Gegenstände der vorliegenden Erfindung sind daher DNA-Sequenzen, welche für die erfindungsgemäßen thymostabilen in vitro-Komplexe bzw. akzessorischen in vitro-Komplexe codieren, sowie entsprechende Vektoren, vorzugsweise Expressionsvektoren.

Die erfindungsgemäßen Vektoren enthalten mindestens das Gen für das Gleitklammerprotein und vorzugsweise mindestens ein Gen für ein Kopplungs- oder/und Elongationsprotein, wie sie jeweils oben definiert sind.

Es ist dabei im Rahmen der Erfindung bevorzugt, wenn der Vektor neben den darin bereits enthaltenen DNA-Sequenzen noch geeignete Restriktionsschnittstellen und ggf. Polylinker zur Insertion weiterer DNA-Sequenzen enthält. Besonders bevorzugt ist es, wenn die räumliche Anordnung von bereits enthaltener DNA-Sequenz und zusätzlicher Insertionsstelle nach Expression zur Ausbildung eines Fusionsproteins führt.

Ebenfalls bevorzugt ist es, wenn der erfindungsgemäße Vektor Promotor- oder/und Operatorbereiche enthält, wobei es besonders bevorzugt ist, wenn derartige Promotor- oder/und Operatorbereiche induzierbar oder reprimierbar sind. Dadurch wird eine Steuerung der Expression in Wirtszellen erheblich vereinfacht und kann besonders effizient gestaltet werden.

Derartige Promotor/Operatorbereiche können in einem Expressionsvektor auch mehrfach vorkommen, so daß eine ggf. unabhängige Expression mehrerer DNA-Sequenzen unter Verwendung nur eines Expressionsvektors ermöglicht wird.

Ein besonders bevorzugter Vektor enthält im Rahmen der Erfindung DNA-Sequenzen codierende für alle Bestandteile eines erfindungsgemäßen in vitro-Komplexes oder akzessorischen in vitro-Komplexes.

Ein weiterer Gegenstand der vorliegenden Erfindung ist eine Wirtszelle, enthaltend einen oder mehrere erfindungsgemäße(n) Vektor(en), wobei in dieser Wirtszelle unter geeigneten Bedingungen die Expression zu Proteinen erfolgen kann. Geeignete Bedingungen schließen beispielsweise Anwesenheit eines Induktors oder eines Derepressors ein.

Für die Transformation, Phageninfektion und Zellkultur existieren Standardprotokolle in Maniatis et al. (Molecular Cloning: A laboratory Manual, Cold Spring Harbor Laboratory Press, N.Y.). Aus der Vielzahl der vorhandenen E. coli Stämme, die zur Transformation geeignet sind, sind die bevorzugten JM101 (ATCC No. 33876), XL1 (Stratagene), RRI (ATCC No. 31343) und BL21 (Pharmacia). Proteinexpression kann z. B. auch geschehen unter Zuhilfenahme des E. coli Stammes INVαF (Invitrogen).

Die Transformanten werden gemäß den geeigneten Wachstumsbedingungen des Wirtsstammes kultiviert. So werden die meisten E. coli Stämme z. B. in LB Medium kultiviert, bei 30°C bis 42°C bis zur logarithmischen oder stationären Wachstumsphase. Die Proteine können aus einer transformierten Kultur gereinigt werden, wobei dies entweder aus einem Zellpellet, nach Zentrifugation, oder aber aus der Kulturlösung geschehen kann. Sofern die Proteine aus dem Zellpellet gereinigt werden, werden die Zellen in einem geeigneten Puffer resuspendiert und mittels Sonifizierung, enzymatischer Behandlung oder Einfrieren und Auftauen aufgebrochen. Sofern die Reinigung aus der Kultursuspension geschieht, das heißt alleine oder über ein Fusionsprotein, wird der Überstand von den Zellen mittels bekannter Verfahren, wie Zentrifugation abgetrennt.

Die Separation und Reinigung der Proteine der erfindungsgemäßen Komplexe entweder aus dem Überstand der Kulturlösung oder aus dem Zellextrakt kann durch bekannte Separations- oder Reinigungsverfahren geschehen. Diese Methoden sind z. B. solche die Löslichkeiten betreffen, wie Salzfallungen, und Lösungsmittelfällungen, Methoden die sich die unterschiedlichen Molekulargewichte zu Nutze machen, wie Dialyse, Ultrafiltration, Gelfiltration, und SDS-Polyacrylamidgelelektrophorese, Methoden die sich die unterschiedlichen Ladungen zu Nutze machen, Ionenaustauschchromatographie, Methoden die sich die unterschiedlichen Hydrophobizitäten zu Nutze machen, wie reverse-phase HPLC (High Performance Liquid Chromatography), Methoden die sich bestimmte Affinitäten zu Nutze machen, wie Affinitätschromatographie, und Methoden die sich Unterschiede im Isoelektrischen Punkt zu Nutze machen, wie isoelektrische Fokussierung. Es ist auch vorstellbar, daß Zellextrakte, entweder aus dem Organismus, welcher das Gen des akzessori-

schen Komplexes trägt, gemacht werden können, welcher die erfindungsgemäße Aufgabe erfüllt, oder aus dem rekombinanten Wirtsorganismus, z. B. E. coli. Mit diesen Extraktionen könnte man unter Umständen andere Reinigungsschritte umgehen.

Die oben beschriebenen Verfahren können in einer Vielzahl von Kombinationen eingesetzt werden um die Proteine des in vitro-Komplexes bereitzustellen.

Der erfindungsgemäße thermostabile prokaryontische akzessorische in vitro Komplexes kann zur Elongation von Nukleinsäuren verwendet werden, z. B. zur Polymerase-Ketten-Reaktion, DNA Sequenzierung, zur Markierung von Nukleinsäuren und anderen Reaktionen, die die Nukleinsäure in vitro Synthese beinhalten. Auch möglich ist die Anwendung in der reversen Transkription, wobei entweder der erfindungsgemäße Komplex selbst reverse Transkriptase Aktivität besitzt, oder aber zusätzlich ein geeignetes Enzym hinzugefügt wird, das reverse Transkriptase Aktivität besitzt.

Ein weiterer Gegenstand der vorliegenden Erfindung ist daher ein Verfahren zur Template-abhängigen Elongation von Nukleinsäuren, wobei die Nukleinsäure nötigenfalls denaturiert, mit mindestens einem Primer unter Hybridisierungsbedingungen versehen wird, wobei der Primer komplementär zu einem flankierenden Bereich einer gewünschten Nukleinsäuresequenz des Templatestrangs ist und mit Hilfe einer Polymerase in Anwesenheit von Nukleotiden eine Primerelongation erfolgt, wobei als Polymerase ein erfindungsgemäßer thermostabiler in vitro-Komplex eingesetzt wird.

Verfahren zur Template-abhängigen Elongation von Nukleinsäuren, bei denen die Elongation ausgehend von einem Primer erfolgt, der an die Template-Nukleinsäure anhybridisiert wurde und ein freies 3'-OH-Ende für die Elongation zur Verfügung stellt, sind dem Fachmann bekannt. Zur Amplifikation wird insbesondere eine Polymerasekettenreaktion durchgeführt. Hierbei wird normalerweise von einer doppelsträngigen DNA-Sequenz ausgegangen, von welcher ein bestimmter Zielbereich amplifiziert werden soll. Hierbei werden zwei Primer eingesetzt, welche zu die Zielsequenz flankierenden Bereichen auf jeweils einem Teilstrang des DNA-Doppelstranges komplementär sind. Zur Anhybridisierung der Primer werden allerdings die DNA-Doppelstränge zuerst denaturiert, insbesondere thermisch aufgeschmolzen. Nach Anhybridisierung der Primer erfolgt eine Elongation mittels der Polymerase, daraufhin wird nochmals denaturiert und damit die neu gebildeten DNA-Stränge von den Template-Strängen getrennt, woraufhin für einen weiteren Elongationszyklus neben den ursprünglichen Templatesträngen auch die im ersten Schritt gebildeten Nukleinsäurestränge als Template zur Verfügung stehen, diese jeweils erneut mit Primern hybridisiert werden und eine erneute Elongation stattfindet. Diese Vorgehensweise wird zyklisch durchgeführt unter jeweils thermischer Denaturierung als Zwischenschritte.

Zur erfindungsgemäß bevorzugten Reversen Transkription von RNA in DNA wird ebenfalls ein erfindungsgemäßer thermostabiler in vitro-Komplex eingesetzt, wobei dessen Elongationsprotein eine Reverse Transkriptase-Aktivität aufweist. Diese Reverse Transkriptase-Aktivität kann die einzige Polymeraseaktivität des Elongationsproteins sein, kann aber auch zusätzlich zu einer vorhandenen 5'-3'-DNA-Polymeraseaktivität vorliegen.

Ein weiteres bevorzugtes erfindungsgemäßes Verfahren betrifft die Sequenzierung von Nukleinsäuren ausgehend von einem Primer, der zu einem der zu sequenzierenden Nukleinsäure benachbarten Bereich komplementär ist, wobei wiederum eine Template-abhängige Elongation oder aber bei Sequenzierung einer RNA eine Reverse Transkription unter Verwendung von Deoxynukleotiden und Dideoxynukleotiden gemäß der Methode von Sanger durchgeführt wird. Als Deoxynukleotide oder Dideoxynukleotide werden im Rahmen dieser bevorzugten Ausführungsform auch die oben beschriebenen jeweiligen Derivate als geeignet angesehen. Insbesondere ist es für die erfindungsgemäßen Verfahren zur Elongation von Nukleinsäuren bevorzugt, daß die gebildeten Nukleinsäuren markiert werden. Hierzu ist es möglich, markierte Primer oder/und markierte Deoxynukleotide oder/und markierte Dideoxynukleotide oder/und markierte Ribonukleotide oder jeweils entsprechende Derivate, wie sie oben bereits beispielsweise beschrieben sind, einzusetzen.

Wiederum ein weiterer Gegenstand der vorliegenden Erfindung ist ein Verfahren zur Markierung von Nukleinsäuren durch Einfügung einzelner Brüche in Phosphodiesterbindungen der Nukleinsäurekette und Ersatz eines Nukleotids an den Bruchstellen durch ein markiertes Nukleotid mit Hilfe einer Polymerase, wobei als Polymerase ein erfindungsgemäßer thermostabiler in vitro-Komplex eingesetzt wird.

Ein solches Verfahren, allgemein Nick-Translation genannt, ermöglicht eine einfache Markierung von Nukleinsäuren. Alle oben bereits beschriebenen markierten Ribonukleotide oder Deoxyribonukleotide oder Derivate davon sind hierfür geeignet, solange die Polymerase sie als Substrat akzeptiert.

Diese erfindungsgemäße Markierung kann z. B. auch im Rahmen oder anschließend an eine PCR-Reaktion erfolgen, wofür die erfindungsgemäße Verwendung eines thermostabilen in vitro-Komplexes besonders günstig ist.

Gegenstand der vorliegenden Erfindung ist ebenfalls ein Kit zur Elongation oder/und Amplifikation oder/und Reversen Transkription oder/und Sequenzierung von Nukleinsäuren, wobei dieser Kit in einem oder in mehreren Behältern vorliegen kann, enthaltend

- a) einen erfindungsgemäßen thermostabilen in vitro Komplex oder
- b) einen thermostabilen, akzessorischen in vitro-Komplex und ggf. separat davon ein Polymeraseaktivität aufweisendes Elongationsprotein sowie ggf. Primer, Puffersubstanzen, Nukleotide, ATP, andere Kofaktoren oder/und Pyrophosphat.

Insbesondere ist Gegenstand der vorliegenden Erfindung ein Kit zur Elongation, Amplifikation, Reversen Transkription, Markierung bzw. Sequenzierung von Nukleinsäuren, wobei zusätzlich Deoxynukleotide bzw. deren Derivate enthaltend sind.

Ein bevorzugter erfindungsgemäßer Reagenzienkit enthält daher zur Amplifikation von Nukleinsäuren neben den Substanzen a) oder b), welche 5'-3'-Polymeraseaktivität aufweisen, auch Deoxynukleotide oder/und Derivate davon. Gegebenenfalls können hier auch Ribonukleotide oder Derivate davon eingesetzt werden, nämlich dann, wenn eine Polymerase eingesetzt wird, welche auch Ribonukleotide als Substrat akzeptiert.

Ein weiterer bevorzugter Gegenstand der vorliegenden Erfindung ist ein Kit zur Sequenzierung von Nukleinsäuren, wobei zusätzlich zu Deoxynukleotiden bzw. deren Derivaten, Dideoxynukleotide bzw. deren Derivate zur Kettentermination enthaltend sind.

Desweiteren ist insbesondere Gegenstand der vorliegenden Erfindung ein Kit zur reversen Transkription von Nukleinsäuren, wobei entweder der erfindungsgemäße Komplex selbst reverse Transkriptase Aktivität besitzt, oder aber zusätzlich ein geeignetes Enzym anwesend ist, das reverse Transkriptase Aktivität besitzt, wobei Deoxynukleotide bzw. deren Derivate im Reaktionsgemisch enthalten sind.

In einer weiteren besonders bevorzugten Ausführungsform enthält der Kit Primer oder/und Deoxynukleotide oder/und Dideoxynukleotide oder/und Ribonukleotide oder/und deren jeweilige Derivate in markierter Form.

Insbesondere für die Sequenzierung von Nukleinsäuren ist es nötig, eine Markierung einzufügen. Geeignete Markierungen sind weiter oben bereits in beispielhafter Form beschrieben und gelten ebenso als bevorzugte Ausführungsformen im Rahmen der erfindungsgemäßen Reagenzienkits.

Im Rahmen der vorliegenden Erfindung ist es weiterhin möglich, daß der Reagenzienkit zur Markierung von Nukleinsäuren im Rahmen einer sogenannten Nick-Translation eingesetzt wird. In diesem Fall enthält der Reagenzienkit die Bestandteile a) oder b) und markierte Nukleotide, wobei Puffersubstanzen, ATP oder andere Kofaktoren oder/und Pyrophosphat ebenfalls anwesend sein können. Primer werden allerdings im Rahmen einer Nick-Translation in der Regel nicht benötigt.

Gegenstand der vorliegenden Erfindung ist desweiteren die Verwendung eines thermostabilen Gleitklammerproteins in *in vitro* Verfahren zur Elongation, Amplifikation, Markierung bzw. Sequenzierung oder Reversen Transkription von Nukleinsäuren.

Bevorzugt ist der Kit, wenn als Gleitklammerprotein ein thermostabiles Homolog zu dem PCNA-Protein (SEQ ID NO: 12, 13, 14, 15), dem PCNA-Homologen aus *Archaeoglobus fulgidus* oder dem  $\beta$ -clamp-Protein (SEQ ID NO: 35, 36) vorliegt. Insbesondere ist ein Kit bevorzugt, der zusätzlich ein Kopplungsprotein enthält.

Desweiteren ist ein Kit bevorzugt, der zusätzlich ein weiteres Protein enthaltend, wobei das weitere Protein als Klammerlader zur Verfügung steht, auch aus der Gruppe der Prokaryonten stammt und thermostabil ist.

Insbesondere ist ein Kit bevorzugt der einen geeigneten Puffer enthält, wie oben beschrieben. Bevorzugt ist ebenfalls, daß der erfindungsgemäße Kit zusätzlich eine Pyrophosphatase, ATP oder/und andere Kofaktoren enthält.

Die folgenden Beispiele sollen in Verbindung mit den Abbildungen die Erfindung näher erläutern:

#### Abbildungen

Im folgenden werden häufig Sequenznamen verwendet, unter denen die Protein- oder Nukleinsäuresequenzen in der Genbank und der EMBL Datenbank stehen.

#### Abb. 1

Proteinsequenzen, paarweise Alignments und multiple Alignments aus Archaeobakterien und den entsprechenden humanen Genen des Replikationsapparates.

Wobei die Annotation <sup>1</sup> bedeutet %Identität zu dem entsprechenden humanen Gen, berechnet aus dem paarweisen Alignment (siehe Anhang) mit BLASTP 2.0.4 [Feb-24-1998] und die Annotation <sup>2</sup> %Identität zu dem entsprechenden Gen von *Archaeoglobus fulgidus*, berechnet aus dem paarweisen Alignment (siehe Anhang) mit BLASTP 2.0.4 [Feb-24-1998] und die Annotation <sup>3</sup> %Identität zu dem entsprechenden humanen Gen, berechnet aus dem paarweisen Alignment mit FASTA 3.1i02 [March, 1998] <sup>5</sup> bedeutet. Die Verfahren werden näher beschrieben in:

Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schaffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", *Nucleic Acids Res.* 25: 3389-3402 und W.R. Pearson & D.J. Lipman *PNAS* (1988) 85: 2444-2448. Die Abbildung zeigt im Falle des Gleitklammerladers I, Gleitklammerlader II, Gleitklammer, koppelnde Untereinheit und Elongationsprotein I die Sequenznamen aus den Datenbanken und ebenso deren SEQ ID Nummer, wobei die Werte in den Klammern jeweils Prozent Identität je Anzahl Aminosäuren darstellen. Im Falle des Elongationsprotein II beziehen sich die Werte auf Prozent Sequenzidentität zu der *Archaeoglobus fulgidus* Sequenz.

#### Abb. 2

Proteinsequenzen, paarweise Alignments und multiple Alignments aus Eubakterien des Replikationsapparates. Wobei die Annotation <sup>1</sup> bedeutet %Identität zu dem entsprechenden Gen aus *E. coli*, berechnet aus dem paarweisen Alignment (siehe Anhang) mit BLASTP 2.0.4 [Feb-24-1998] <sup>5</sup>. Das Verfahren wird in: Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schaffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", *Nucleic Acids Res.* 25: 3389-3402 näher beschrieben.

#### Abb. 3

Skizze des Replikationsapparates in einer möglichen Variante, wobei die Gleitkammer über eine koppelnde Untereinheit an das Elongationsprotein bindet.

#### Abb. 4

Die Abb. 4 zeigt Alignments zweier konservierter Regionen des Gleitklammerproteins, sowie die daraus hergeleiteten Konsensussequenzen. Folgende Gene sind gezeigt: PCNA human (aus SEQ ID NO: 11), die entsprechende Sequenz aus *Archaeoglobus fulgidus* (aus SEQ ID NO: 12), aus *Methanococcus janashii* (aus SEQ ID NO: 13), aus *Pyrococcus horikoshii* (aus SEQ ID NO: 14) und aus *Methanococcus thermoautotrophicus* (aus SEQ ID NO: 15).

## Abb. 5

Die Abb. 5 zeigt ein Alignment einer konservierten Regionen der koppelnden Untereinheit, sowie die daraus hergeleiteten Konsensussequenzen. Folgende Gene sind gezeigt: PfuORF2, DPD2\_HUMAN, AF1790 und MJ0702. Die SEQ ID Nummern können Abb. 1 entnommen werden.

## Abb. 6

Die Abb. 6 zeigt ein Alignment einer konservierten Regionen der koppelnden Untereinheit, sowie die daraus hergeleiteten Konsensussequenzen. Folgende Gene sind gezeigt: AC11\_HUMAN, AF2060, MTH0241, PHBN012 und MJ1422. Die SEQ ID Nummern können Abb. 1 entnommen werden.

## Abb. 7

Die Abb. 7 zeigt ein Alignment einer konservierten Regionen des Gleitklammerlader 2, sowie die daraus hergeleiteten Konsensussequenzen. Folgende Gene sind gezeigt: AC15\_HUMAN, MJ0884, AF1195, MTH0240 und MTH0240. Die SEQ ID Nummern können Abb. 1 entnommen werden.

## Abb. 8

Die Abb. 8 zeigt ein Alignment einer konservierten Regionen des Elongationsproteins 1, sowie die daraus hergeleiteten Konsensussequenzen. Folgende Gene sind gezeigt: DPOD\_HUMAN, MJ0885, MTH1208, PHBT047 und DPOL\_ARCFU. Die SEQ ID Nummern können Abb. 1 entnommen werden.

## Abb. 9

Die Abb. 9 zeigt ein Alignment einer konservierten Regionen des Elongationsproteins 2, sowie die daraus hergeleiteten Konsensussequenzen. Folgende Gene sind gezeigt: AF1722, MJ1630, PfuORF3, MTH1536 und PHBN021. Die SEQ ID Nummern können Abb. 1 entnommen werden.

## Abb. 10

Die Abb. 10 zeigt ein Alignment einer konservierten Regionen des Elongationsproteins aus Eubakterien, sowie die daraus hergeleiteten Konsensussequenzen. Folgende Gene sind gezeigt: DP3A\_ECOLI: DNA Pol III, alpha subunit, Escherichia coli, BB0579: DNA Pol III, alpha subunit, Borrelia burgdorferi, DP3A\_HELPY: DNA Pol III, alpha subunit, Helicobacter pylori AA50: Aquifex aeolicus, section 50 und DP3A\_SALTY: DNA Pol III, alpha subunit, Salmonella typhimurium).

## Abb. 11

Die Abb. 11 zeigt ein Alignment einer konservierten Regionen der Gleitklammer aus Eubakterien, sowie die daraus hergeleiteten Konsensussequenzen. Folgende Gene sind gezeigt: AAPOL3B, DP3B\_ECOLI, S.TYPHIM, DP3B\_PROMI, DP3B\_PSEPU und DP3B\_STRCO (AAPOL3B: Aquifex Aeolicus sektion 93: DP3B\_ECOLI: DNA Pol III, beta chain, Escherichia coli S.TYPHIM: DNA Pol III, beta chain, Salmonella typhimurium P3B\_PROMI: DNA Pol III, beta chain, Probeus mirabilis DP3B\_PSEPU: DNA Pol III, beta chain, Pseudomonas putida DP3B\_STRCO: DNA Pol III, beta chain, Strptomyces coelicolor).

## Abb. 12

Multiples Alignment der Gleitklammer-proteinsequenzen zur Generierung des Hidden Markov Models.

## Abb. 13

Multiples Alignment der eubakteriellen Gleitklammerproteinsequenzen zur Generierung des Hidden Markov Models (AAPOL3B: Aquifex Aeolicus sektion 93: DP3B\_ECOLI: DNA Pol III, beta chain, Escherichia coli S.TYPHIM: DNA Pol III, beta chain, Salmonella typhimurium P3B\_PROMI: DNA Pol III, beta chain, Probeus mirabilis DP3B\_PSEPU: DNA Pol III, beta chain, Pseudomonas putida DP3B\_STRCO: DNA Pol III, beta chain, Strptomyces coelicolor).

## Abb. 14

Multiples Alignment der Gleitklammerlader 1 Proteinsequenzen zur Generierung des Hidden Markov Models.

## Abb. 15

Multiples Alignment der Gleitklammerlader 2 Proteinsequenzen zur Generierung des Hidden Markov Models.

## Abb. 16

Multiples Alignment der Proteinsequenzen der koppelnden Untereinheiten zur Generierung des Hidden Markov Models.

## Abb. 17

Multiples Alignment der Sequenzen der Elongationsproteine 1 zur Generierung des Hidden Markov Models.

## Abb. 18

Multiples Alignment der Sequenzen der Elongationsproteine 2 zur Generierung des Hidden Markov Models.

## Abb. 19

Multiples Alignment der Sequenzen der eubakteriellen Elongationsproteine zur Generierung des Hidden Markov Models. Folgende Gene sind gezeigt:

DP3A\_ECOLI: DNA Pol III, alpha subunit, Escherichia coli, BB0579: DNA Pol III, alpha subunit, Borrelia burgdorferi, DP3A\_HELPY: DNA Pol III, alpha subunit, Helicobacter pylori AA50: Aquifex aeolicus, section 50 und DP3A\_SALTY: DNA Pol III, alpha subunit, Salmonella typhimurium).

## Beispiel

Die DNA wird mittels der bekannten Verfahren aus dem Organismus *Archaeoglobus fulgidus* (DSM No. 4304) gereinigt. Organismen Aufzucht geschieht hierbei durch die DSM (Deutsche Sammlung für Mikroorganismen). Um die entsprechenden Gene (Gleitklammerlader I/II, Gleitklammer, Elongationsproteine I/II, koppelnde Untereinheit) in den Expressionsvektor pTrec99 zu klonieren, werden für jedes Gen Primer entwickelt, die den vollständigen offenen Leserahmen umspannen, sowie Start- wie Stopcodon. Hierbei werden den Primern jeweils Restriktionsenden hinzugefügt, die die gerichtete Klonierung in den Expressionsvektor erleichtern. Unter Verwendung von etwa 200 ng Gesamtgenomischer DNA werden mit den entsprechenden Annealing-Temperaturen PCR Reaktionen gefahren (etwa 35 Zyklen) und die daraus resultierenden Produkte gereinigt. Nach der Reinigung werden die Produkte mit Restriktionsenzymen behandelt und über ein Agarosegel gereinigt um zur Ligation bereit zu stehen. Der Expressionsvektor wird mittels Restriktionsenzymen so linearisiert, gereinigt und verdünnt, daß er zur Ligation mit den Amplifikaten der akzessorischen Gene aus der obigen PCR bereit ist. Die Ligation wird angesetzt und nach Inkubation ein Aliquot in den *E. coli* Stamm INVα-haF<sup>+</sup> (Invitrogen) transformiert. Von jedem Gen werden 3 positive Kolonien gepickt, Plasmid DNA präpariert und die Inserts auf Vollständigkeit und Richtigkeit mittels DNA Sequenzierung überprüft. Korrekte Klone werden ausgesucht und erneut zur Vereinzelnung auf Agarplatten (ampicillin) ausgebracht. Kolonien werden gepickt und Übermächtkulturen angesetzt. Ein Aliquot (500 µl) der Übermächtkultur wird in eine ein bis fünf Liter Kultur von LB (Ampicillin: 80 mg/l) gegeben. Die Kulturen wachsen bis zu einer OD<sub>660</sub> von 0.8 bei 37°C nun wird zur Induktion IPTG zugegeben (125 mg/l). Diese Kulturen wachsen nun weitere 11 Stunden. Die Kulturen werden zentrifugiert und die Pellets in einem Puffer aufgenommen (Puffer A: 50 mM Tris-HCl pH 7.9, 50 mM Dextrose, 1 mM EDTA). Nach Zentrifugation werden die Zellen erneut aufgenommen, jedoch enthält Puffer A nun zusätzlich Lysozym (4 mg/ml). Nach Inkubation (15 min.) wird ein gleiches Volumen Puffer B zugegeben (B: 10 mM Tris-HCl pH 7.9, 50 mM KCl, 1 mM EDTA, 1 mM PMSF, 0.5% Tween 20, 0.5% Nonidet P40) und die Lyse zur Inkubation bei 75°C für eine Stunde gegeben. Nach Zentrifugation wird der Überstand entnommen und die überexprimierten Proteine werden mittels (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> ausgefällt. Die Pellets werden nach Zentrifugation gesammelt und die Proteine mit Puffer A resuspendiert. Die resuspendierten Proteine werden gegen Aufbewahrungspuffer (50 mM Tris-HCl pH 7.9, 50 mM KCl, 0.1 mM EDTA, 1 mM DTT, 0.5 mM PMSF, 50% Glycerol) dialysiert und anschließend bei -70°C gelagert.

Um die Aktivität der Proteine zu testen werden Reaktionen wie folgt zusammengesetzt: Aliquote der Proteine werden in unterschiedlichen Konfigurationen und Molaritäten vereint, Gleitklammerlader I/II mit Gleitklammer, koppelnder Untereinheit und Elongationsprotein I, oder Gleitklammerlader I/II mit Gleitklammer, mit und ohne koppelnder Untereinheit und Elongationsprotein II, oder Gleitklammer, und Elongationsprotein I oder II und schließlich nur Elongationsprotein I oder II; als Puffer dient der obige Aufbewahrungspuffer. DNA Polymerisationsaktivität wird mittels Inkorporation von (Methyl-<sup>3</sup>H) TTP in Trichlorsäure-unlösliches Material gemessen (Ishino, Y., Iwasaki, H., Fukui, H., Mineno, J., Kato, I., & Schinigawa, H. (1992) *Biochimie* 74, 131-136). Um die Prozessivität zu bestimmen werden die obigen Proteingemische in Primer-Elongationsexperimenten verwendet. Ein M13 Einzelstrangtemplate wird in 10 mM Tris-HCl (pH 9.4) eingebracht und gemeinsam mit einem Universal-primer (5'-FITC markiert) erhitzt (92°C) und abgekühlt (Raumtemperatur). Verdünnungsserien des so generierten Template-Primer Gemisches werden in einer Reaktion bestehend aus Nukleotiden (etwa 200 µM bis 1 mM), Reaktionspuffer (Endkonzentration: 50 mM KCl, 10 mM Tris-HCl (pH 8.3), 1.5-5 mM MgCl<sub>2</sub>, ATP (0 mM-200 mM) und proteinstabilisierenden Agenzien zusammengeführt und für 10 Minuten bei 37°C, 52°, 62°, 68°, 74°, und 78° inkubiert. Ein Aliquot wird zur Analyse auf einen Sequenzautomaten geladen (z. B. Alf, Pharmacia Biotech). Die obigen Proteingemische dienen auch dazu Fidelity und Exonukleaseaktivität zu messen, wobei die in Kohler et al. (*Proc. Natl. Acad. Sci. USA* 88: 7958-7962 (1991) oder Chase et al. (*J. Biol. Chem.*, 249: 4545-4552 (1972) beschriebenen Verfahren zur Anwendung kommen. Ebenso werden die Proteingemische in der PCR eingesetzt (*Methods in Molecular Biology* Vol. 15 Humana Press Totowa, New Jersey, 1993, edited by Bruce A. White).

## (1) ALLGEMEINE ANGABEN:

## (i) ANMELDER:

- (A) NAME: LION bioscience AG  
(B) STRASSE: Im Neuenheimer Feld 517  
(C) ORT: Heidelberg  
(E) LAND: DE  
(F) POSTLEITZAHL: 69120

(ii) BEZEICHNUNG DER ERFINDUNG: Akzessorische Komplexe mit  
Polymeraseaktivitaet

(iii) ANZAHL DER SEQUENZEN: 38

## (iv) COMPUTER-LESBARE FASSUNG:

- (A) DATENTRÄGER: Floppy disk  
(B) COMPUTER: IBM PC compatible  
(C) BETRIEBSSYSTEM: PC-DOS/MS-DOS  
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPA)

## (2) ANGABEN ZU SEQ ID NO: 1:

## (i) SEQUENZKENNZEICHEN:

- (A) LÄNGE: 340 Aminosäuren  
(B) ART: Aminosäure  
(C) STRANGFORM: Einzelstrang  
(D) TOPOLOGIE: beides

(ii) ART DES MOLEKÜLS: Protein

## (vi) URSPRÜNGLICHE HERKUNFT:

- (A) ORGANISMUS: Homo sapiens

## (xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 1:

Met Glu Thr Ser Ala Leu Lys Gln Gln Glu Gln Pro Ala Ala Thr Lys  
1 5 10 15

Ile Arg Asn Leu Pro Trp Val Glu Lys Tyr Arg Pro Gln Thr Leu Asn  
20 25 30



# DE 198 40 771 A 1

Asp Leu Ile Ser His Gln Asp Ile Leu Ser Thr Ile Gln Lys Phe Ile

35 40 45

Asn Glu Asp Arg Leu Pro His Leu Leu Leu Tyr Gly Pro Pro Gly Thr

50 55 60

Gly Lys Thr Ser Thr Ile Leu Ala Cys Ala Lys Gln Leu Tyr Lys Asp

65 70 75 80

Lys Glu Phe Gly Ser Met Val Leu Glu Leu Asn Ala Ser Asp Asp Arg

85 90 95

Gly Ile Asp Ile Ile Arg Gly Pro Ile Leu Ser Phe Ala Ser Thr Arg

100 105 110

Thr Ile Phe Lys Lys Gly Phe Lys Leu Val Ile Leu Asp Glu Ala Asp

115 120 125

Ala Met Thr Gln Asp Ala Gln Asn Ala Leu Arg Arg Val Ile Glu Lys

130 135 140

Phe Thr Glu Asn Thr Arg Phe Cys Leu Ile Cys Asn Tyr Leu Ser Lys

145 150 155 160

Ile Ile Pro Ala Leu Gln Ser Arg Cys Thr Arg Phe Arg Phe Gly Pro

165 170 175

Leu Thr Pro Glu Leu Met Val Pro Arg Leu Glu His Val Val Glu Glu

180 185 190

Glu Lys Val Asp Ile Ser Glu Asp Gly Met Lys Ala Leu Val Thr Leu

195 200 205

Ser Ser Gly Asp Met Arg Arg Ala Leu Asn Ile Leu Gln Ser Thr Asn

210 215 220

Met Ala Phe Gly Lys Val Thr Glu Glu Thr Val Tyr Thr Cys Thr Gly

225 230 235 240

His Pro Leu Lys Ser Asp Ile Ala Asn Ile Leu Asp Trp Met Leu Asn

245 250 255

Gln Asp Phe Thr Thr Ala Tyr Arg Asn Ile Thr Glu Leu Lys Thr Leu

260 265 270

Lys Gly Leu Ala Leu His Asp Ile Leu Thr Glu Ile His Leu Phe Val

275 280 285

# DE 198 40 771 A 1

His Arg Val Asp Phe Pro Ser Ser Val Arg Ile His Leu Leu Thr Lys  
290 295 300

Met Ala Asp Ile Glu Tyr Arg Leu Ser Val Gly Thr Asn Glu Lys Ile  
305 310 315 320

Gln Leu Ser Ser Leu Ile Ala Ala Phe Gln Val Thr Arg Asp Leu Ile  
325 330 335

Val Ala Glu Ala  
340

## (2) ANGABEN ZU SEQ ID NO: 2:

### (i) SEQUENZKENNZEICHEN:

(A) LÄNGE: 319 Aminosäuren

(B) ART: Aminosäure

(C) STRANGFORM: Einzelstrang

(D) TOPOLOGIE: linear

### (ii) ART DES MOLEKÜLS: Protein

### (vi) URSPRÜNGLICHE HERKUNFT:

(A) ORGANISMUS: Archaeoglobus fulgidus

### (xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 2:

Met Glu Asn Phe Glu Ile Trp Val Glu Lys Tyr Arg Pro Arg Thr Leu  
1 5 10 15

Asp Glu Val Val Gly Gln Asp Glu Val Ile Gln Arg Leu Lys Gly Tyr  
20 25 30

Val Glu Arg Lys Asn Ile Pro His Leu Leu Phe Ser Gly Pro Pro Gly  
35 40 45

Thr Gly Lys Thr Ala Thr Ala Ile Ala Leu Ala Arg Asp Leu Phe Gly  
50 55 60

Glu Asn Trp Arg Asp Asn Phe Ile Glu Met Asn Ala Ser Asp Glu Arg  
65 70 75 80

Gly Ile Asp Val Val Arg His Lys Ile Lys Glu Phe Ala Arg Thr Ala  
85 90 95

# DE 198 40 771 A 1

Pro Ile Gly Gly Ala Pro Phe Lys Ile Ile Phe Leu Asp Glu Ala Asp  
100 105 110

Ala Leu Thr Ala Asp Ala Gln Ala Ala Leu Arg Arg Thr Met Glu Met  
115 120 125

Tyr Ser Lys Ser Cys Arg Phe Ile Leu Ser Cys Asn Tyr Val Ser Arg  
130 135 140

Ile Ile Glu Pro Ile Gln Ser Arg Cys Ala Val Phe Arg Phe Lys Pro  
145 150 155 160

Val Pro Lys Glu Ala Met Lys Lys Arg Leu Leu Glu Ile Cys Glu Lys  
165 170 175

Glu Gly Val Lys Ile Thr Glu Asp Gly Leu Glu Ala Leu Ile Tyr Ile  
180 185 190

Ser Gly Gly Asp Phe Arg Lys Ala Ile Asn Ala Leu Gln Gly Ala Ala  
195 200 205

Ala Ile Gly Glu Val Val Asp Ala Asp Thr Ile Tyr Gln Ile Thr Ala  
210 215 220

Thr Ala Arg Pro Glu Glu Met Thr Glu Leu Ile Gln Thr Ala Leu Lys  
225 230 235 240

Gly Asn Phe Met Glu Ala Arg Glu Leu Leu Asp Arg Leu Met Val Glu  
245 250 255

Tyr Gly Met Ser Gly Glu Asp Ile Val Ala Gln Leu Phe Arg Glu Ile  
260 265 270

Ile Ser Met Pro Ile Lys Asp Ser Leu Lys Val Gln Leu Ile Asp Lys  
275 280 285

Leu Gly Glu Val Asp Phe Arg Leu Thr Glu Gly Ala Asn Glu Arg Ile  
290 295 300

Gln Leu Asp Ala Tyr Leu Ala Tyr Leu Ser Thr Leu Ala Lys Lys  
305 310 315

(2) ANGABEN ZU SEQ ID NO: 3:

(i) SEQUENZKENNZEICHEN:

(A) LÄNGE: 1847 Aminosäuren

(B) ART: Aminosäure

(C) STRANGFORM: Einzelstrang

(D) TOPOLOGIE: linear

5 (ii) ART DES MOLEKÜLS: Protein

(vi) URSPRÜNGLICHE HERKUNFT:

10 (A) ORGANISMUS: Methanococcus jannaschii

15 (xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 3:

Met Val Ile Ile Met Glu Lys Pro Trp Val Glu Lys Tyr Arg Pro Lys  
 1 5 10 15

20 Thr Leu Asp Asp Ile Val Gly Gln Asp Glu Ile Val Lys Arg Leu Lys  
 20 25 30

25 Lys Tyr Val Glu Lys Lys Ser Met Pro His Leu Leu Phe Ser Gly Pro  
 35 40 45

30 Pro Gly Val Gly Lys Cys Leu Thr Gly Asp Thr Lys Val Ile Val Asn  
 50 55 60

35 Gly Glu Ile Arg Glu Ile Gly Glu Val Ile Glu Glu Ile Ser Asn Gly  
 65 70 75 80

Lys Phe Gly Val Thr Leu Thr Asn Asn Leu Lys Val Leu Gly Ile Asp  
 85 90 95

40 Glu Asp Gly Lys Ile Arg Glu Phe Asp Val Gln Tyr Val Tyr Lys Asp  
 100 105 110

45 Lys Thr Asn Thr Leu Ile Lys Ile Lys Thr Lys Met Gly Arg Glu Leu  
 115 120 125

50 Lys Val Thr Thr Tyr His Pro Leu Leu Ile Asn His Lys Asn Gly Glu  
 130 135 140

55 Ile Lys Trp Glu Lys Ala Glu Asn Leu Lys Val Gly Asp Lys Leu Ala  
 145 150 155 160

Thr Pro Arg Tyr Ile Leu Phe Asn Glu Ser Asp Tyr Asn Glu Glu Leu  
 165 170 175

60 Ala Glu Trp Leu Gly Tyr Phe Ile Gly Asp Gly His Ala Asp Lys Glu  
 180 185 190

65

# DE 198 40 771 A 1

Ser Asn Lys Ile Thr Phe Thr Asn Gly Asp Glu Lys Leu Arg Lys Arg  
195 200 205

Phe Ala Glu Leu Thr Glu Lys Leu Phe Lys Asp Ala Lys Ile Lys Glu  
210 215 220

Arg Ile His Lys Asp Arg Thr Pro Asp Ile Tyr Val Asn Ser Lys Glu  
225 230 235 240

Ala Val Glu Phe Ile Asp Lys Leu Gly Leu Arg Gly Lys Lys Ala Asp  
245 250 255

Lys Val Arg Ile Pro Lys Glu Ile Met Arg Ser Asp Ala Leu Arg Ala  
260 265 270

Phe Leu Arg Ala Tyr Phe Asp Cys Asp Gly Gly Ile Glu Lys His Ser  
275 280 285

Ile Val Leu Ser Thr Ala Ser Lys Glu Met Ala Glu Asp Leu Val Tyr  
290 295 300

Ala Leu Leu Arg Phe Gly Ile Ile Ala Lys Leu Arg Glu Lys Val Asn  
305 310 315 320

Lys Asn Asn Asn Lys Val Tyr Tyr His Ile Val Ile Ser Asn Ser Ser  
325 330 335

Asn Leu Arg Thr Phe Leu Asp Asn Ile Gly Phe Ser Gln Glu Arg Lys  
340 345 350

Leu Lys Lys Leu Leu Glu Ile Ile Lys Asp Glu Asn Pro Asn Leu Asp  
355 360 365

Val Ile Thr Ile Asp Lys Glu Lys Ile Arg Tyr Ile Arg Asp Arg Leu  
370 375 380

Lys Val Lys Leu Thr Arg Asp Ile Glu Lys Asp Asn Trp Ser Tyr Asn  
385 390 395 400

Lys Cys Arg Lys Ile Thr Gln Glu Leu Leu Lys Glu Ile Tyr Tyr Arg  
405 410 415

Leu Glu Glu Leu Lys Glu Ile Glu Lys Ala Leu Glu Glu Asn Ile Leu  
420 425 430

Ile Asp Trp Asp Glu Val Ala Glu Arg Arg Lys Glu Ile Ala Glu Lys  
435 440 445

5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60  
65

# DE 198 40 771 A 1

Thr Gly Ile Arg Ser Asp Arg Ile Leu Glu Tyr Ile Arg Gly Lys Arg  
 450 455 460

5 Lys Pro Ser Leu Lys Asn Tyr Ile Lys Ile Ala Asn Thr Leu Gly Lys  
 465 470 475 480

10 Asn Ile Glu Lys Ile Ile Asp Ala Met Arg Ile Phe Ala Lys Lys Tyr  
 485 490 495

15 Ser Ser Tyr Ala Glu Ile Gly Lys Met Leu Asn Met Trp Asn Ser Ser  
 500 505 510

20 Ile Lys Ile Tyr Leu Glu Ser Asn Thr Gln Glu Ile Glu Lys Leu Glu  
 515 520 525

25 Glu Ile Arg Lys Thr Glu Leu Lys Leu Val Lys Glu Ile Leu Asn Asp  
 530 535 540

30 Glu Lys Leu Ile Asp Ser Ile Gly Tyr Val Leu Phe Leu Ala Ser Asn  
 545 550 555 560

35 Glu Ile Tyr Trp Asp Glu Ile Val Glu Ile Glu Gln Leu Asn Gly Glu  
 565 570 575

40 Phe Thr Ile Tyr Asp Leu His Val Pro Arg Tyr His Asn Phe Ile Gly  
 580 585 590

45 Gly Asn Leu Pro Thr Ile Leu His Asn Thr Thr Ala Ala Leu Cys Leu  
 595 600 605

50 Ala Arg Asp Leu Phe Gly Glu Asn Trp Arg Asp Asn Phe Leu Glu Leu  
 610 615 620

55 Asn Ala Ser Val Ser Lys Asp Thr Pro Ile Leu Val Lys Ile Asp Gly  
 625 630 635 640

60 Lys Val Lys Arg Thr Thr Phe Glu Glu Leu Asp Lys Ile Tyr Phe Glu  
 645 650 655

65 Thr Asn Asp Glu Asn Glu Met Tyr Lys Lys Val Asp Asn Leu Glu Val  
 660 665 670

70 Leu Thr Val Asp Glu Asn Phe Arg Val Arg Trp Arg Lys Val Ser Thr  
 675 680 685

75 Ile Ile Arg His Lys Val Asp Lys Ile Leu Arg Ile Lys Phe Glu Gly  
 690 695 700

# DE 198 40 771 A 1

Gly Tyr Ile Glu Leu Thr Gly Asn His Ser Ile Met Met Leu Asp Glu  
705 710 715 720

Asn Gly Leu Val Ala Lys Lys Ala Ser Asp Ile Lys Val Gly Asp Cys  
725 730 735

Phe Leu Ser Phe Val Ala Asn Ile Glu Gly Glu Lys Asp Arg Leu Asp  
740 745 750

Leu Lys Glu Phe Glu Pro Lys Asp Ile Thr Ser Arg Val Lys Ile Ile  
755 760 765

Asn Asp Phe Asp Ile Asp Glu Asp Thr Ala Trp Met Leu Gly Leu Tyr  
770 775 780

Val Ala Glu Gly Ala Val Gly Phe Lys Gly Lys Thr Ser Gly Gln Val  
785 790 795 800

Ile Tyr Thr Leu Gly Ser His Glu His Asp Leu Ile Asn Lys Leu Asn  
805 810 815

Asp Ile Val Asp Lys Lys Gly Phe Ser Lys Tyr Glu Asn Phe Thr Gly  
820 825 830

Ser Gly Phe Asp Arg Lys Arg Leu Ser Ala Lys Gln Ile Arg Ile Leu  
835 840 845

Asn Thr Gln Leu Ala Arg Phe Val Glu Glu Asn Phe Tyr Asp Gly Asn  
850 855 860

Gly Arg Arg Ala Arg Asn Lys Arg Ile Pro Asp Ile Ile Phe Glu Leu  
865 870 875 880

Lys Glu Asn Leu Arg Val Glu Phe Leu Lys Gly Leu Ala Asp Gly Asp  
885 890 895

Ser Ser Gly Asn Trp Arg Glu Val Val Arg Ile Ser Ser Lys Ser Asp  
900 905 910

Asn Leu Leu Ile Asp Thr Val Trp Leu Ala Arg Ile Ser Gly Ile Glu  
915 920 925

Ser Ser Ile Phe Glu Asn Glu Ala Arg Leu Ile Trp Lys Gly Gly Met  
930 935 940

Lys Trp Lys Lys Ser Asn Leu Leu Pro Ala Glu Pro Ile Ile Lys Met  
945 950 955 960

# DE 198 40 771 A 1

Ile Lys Lys Leu Glu Asn Lys Ile Asn Gly Asn Trp Arg Tyr Ile Leu  
965 970 975

5 Arg His Gln Leu Tyr Glu Gly Lys Lys Arg Val Ser Lys Asp Lys Ile  
980 985 990

10 Lys Gln Ile Leu Glu Met Val Asn Val Glu Lys Leu Ser Asp Lys Glu  
995 1000 1005

15 Lys Glu Val Tyr Asp Leu Leu Lys Lys Leu Ser Lys Thr Glu Leu Tyr  
1010 1015 1020

20 Ala Leu Val Val Lys Glu Ile Glu Ile Ile Asp Tyr Asn Asp Phe Val  
1025 1030 1035 1040

Tyr Asp Val Ser Val Pro Asn Asn Glu Met Phe Phe Ala Gly Asn Val  
1045 1050 1055

25 Pro Ile Leu Leu His Asn Ser Asp Glu Arg Gly Ile Asp Val Ile Arg  
1060 1065 1070

30 Thr Lys Val Lys Asp Phe Ala Arg Thr Lys Pro Ile Gly Asp Val Pro  
1075 1080 1085

35 Phe Lys Ile Ile Phe Leu Asp Glu Ser Asp Ala Leu Thr Ala Asp Ala  
1090 1095 1100

Gln Asn Ala Leu Arg Arg Thr Met Glu Lys Tyr Ser Asp Val Cys Arg  
1105 1110 1115 1120

40 Phe Ile Leu Ser Cys Leu Thr Gly Asp Ala Lys Ile Thr Leu Pro Asp  
1125 1130 1135

45 Glu Arg Glu Ile Lys Ile Glu Asp Phe Ile Lys Met Phe Glu Glu Arg  
1140 1145 1150

50 Lys Leu Lys His Val Leu Asn Arg Asn Gly Glu Asp Leu Val Leu Ala  
1155 1160 1165

Gly Val Lys Phe Asn Ser Lys Ile Val Asn His Lys Val Tyr Arg Leu  
1170 1175 1180

55 Val Leu Glu Ser Gly Arg Glu Ile Glu Ala Thr Gly Asp His Lys Phe  
1185 1190 1195 1200

60 Leu Thr Arg Asp Gly Trp Lys Glu Val Tyr Glu Leu Lys Glu Asp Asp  
1205 1210 1215

65



# DE 198 40 771 A 1

Glu Val Leu Val Tyr Pro Ala Leu Glu Gly Val Gly Phe Glu Val Asp  
 1220 1225 1230  
 Glu Arg Arg Ile Ile Gly Leu Asn Glu Phe Tyr Glu Phe Leu Thr Asn  
 1235 1240 1245  
 Tyr Glu Ile Lys Leu Gly Tyr Lys Pro Leu Gly Lys Ala Lys Ser Tyr  
 1250 1255 1260  
 Lys Glu Leu Ile Thr Arg Asp Lys Glu Lys Ile Leu Ser Arg Val Leu  
 1265 1270 1275 1280  
 Glu Leu Ser Asp Lys Tyr Ser Lys Ser Glu Ile Arg Arg Lys Ile Glu  
 1285 1290 1295  
 Glu Glu Phe Gly Ile Lys Ile Ser Leu Thr Thr Ile Lys Asn Leu Ile  
 1300 1305 1310  
 Asn Gly Lys Ile Asp Gly Phe Ala Leu Lys Tyr Val Arg Lys Ile Lys  
 1315 1320 1325  
 Glu Leu Gly Trp Asp Glu Ile Thr Tyr Asp Asp Glu Lys Ala Gly Ile  
 1330 1335 1340  
 Phe Ala Arg Leu Leu Gly Phe Ile Ile Gly Asp Gly His Leu Ser Lys  
 1345 1350 1355 1360  
 Ser Lys Glu Gly Arg Ile Leu Ile Thr Ala Thr Ile Asn Glu Leu Glu  
 1365 1370 1375  
 Gly Ile Lys Lys Asp Leu Glu Lys Leu Gly Ile Lys Ala Ser Asn Ile  
 1380 1385 1390  
 Ile Glu Lys Asp Ile Glu His Lys Leu Asp Gly Arg Glu Ile Lys Gly  
 1395 1400 1405  
 Lys Thr Ser Phe Ile Tyr Ile Asn Asn Lys Ala Phe Tyr Leu Leu Leu  
 1410 1415 1420  
 Asn Phe Trp Gly Val Glu Ile Gly Asn Lys Thr Ile Asn Gly Tyr Asn  
 1425 1430 1435 1440  
 Ile Pro Lys Trp Ile Lys Tyr Gly Asn Lys Phe Val Lys Arg Glu Phe  
 1445 1450 1455  
 Leu Arg Gly Leu Phe Gly Ala Asp Gly Thr Lys Pro Tyr Ile Lys Lys  
 1460 1465 1470

5  
 10  
 15  
 20  
 25  
 30  
 35  
 40  
 45  
 50  
 55  
 60  
 65

# DE 198 40 771 A 1

Tyr Asn Ile Asn Gly Ile Lys Leu Gly Ile Arg Val Glu Asn Ile Ser  
 1475 1480 1485

5 Lys Asp Lys Thr Leu Glu Phe Phe Glu Glu Val Lys Lys Met Leu Glu  
 1490 1495 1500

10 Glu Phe Glu Val Glu Ser Tyr Ile Lys Val Ser Lys Ile Asp Asn Lys  
 1505 1510 1515 1520

Asn Leu Thr Glu Leu Ile Val Lys Ala Asn Asn Lys Asn Tyr Leu Lys  
 15 1525 1530 1535

Tyr Leu Ser Arg Ile Ser Tyr Ala Tyr Glu Lys Asp Asn Phe Ala Arg  
 20 1540 1545 1550

Leu Val Gly Glu Tyr Leu Arg Ile Lys Glu Ala Tyr Lys Asp Ile Ile  
 25 1555 1560 1565

Leu Lys Glu Ile Ala Glu Asn Ala Leu Lys Glu Ala Asp Gly Glu Lys  
 1570 1575 1580

30 Ser Leu Arg Glu Leu Ala Arg Lys Tyr Asn Val Pro Val Asp Phe Ile  
 1585 1590 1595 1600

Ile Asn Gln Leu Lys Gly Lys Asp Ile Gly Leu Pro Arg Asn Phe Met  
 35 1605 1610 1615

Thr Phe Glu Glu Phe Leu Lys Glu Lys Val Val Asp Gly Lys Tyr Val  
 40 1620 1625 1630

Ser Glu Arg Ile Ile Lys Lys Glu Cys Ile Gly Tyr Arg Asp Val Tyr  
 1635 1640 1645

45 Asp Ile Thr Cys His Lys Asp Pro Ser Phe Ile Ala Asn Gly Phe Val  
 1650 1655 1660

Ser His Asn Cys Asn Tyr Pro Ser Lys Ile Ile Pro Pro Ile Gln Ser  
 50 1665 1670 1675 1680

Arg Cys Ala Val Phe Arg Phe Ser Pro Leu Lys Lys Glu Asp Ile Ala  
 55 1685 1690 1695

Lys Lys Leu Lys Glu Ile Ala Glu Lys Glu Gly Leu Asn Leu Thr Glu  
 1700 1705 1710

60 Ser Gly Leu Glu Ala Ile Ile Tyr Val Ser Glu Gly Asp Met Arg Lys  
 1715 1720 1725

65

# DE 198 40 771 A 1

Ala Ile Asn Val Leu Gln Thr Ala Ala Ala Leu Ser Asp Val Ile Asp  
1730 1735 1740

Asp Glu Ile Val Tyr Lys Val Ser Ser Arg Ala Arg Pro Glu Glu Val  
1745 1750 1755 1760

Lys Lys Met Met Glu Leu Ala Leu Asp Gly Lys Phe Met Ser Glu Ala Arg  
1765 1770 1775

Asp Leu Leu Tyr Lys Leu Met Val Glu Trp Gly Met Ser Gly Glu Asp  
1780 1785 1790

Ile Leu Asn Gln Met Phe Arg Glu Ile Asn Ser Leu Asp Ile Asp Glu  
1795 1800 1805

Arg Lys Lys Val Glu Leu Ala Asp Ala Ile Gly Glu Thr Asp Phe Arg  
1810 1815 1820

Ile Val Glu Gly Ala Asn Glu Arg Ile Gln Leu Ser Ala Leu Leu Ala  
1825 1830 1835 1840

Lys Met Ala Leu Met Gly Arg  
1845

## (2) ANGABEN ZU SEQ ID NO: 4:

### (i) SEQUENZKENNZEICHEN:

- (A) LÄNGE: 855 Aminosäuren
- (B) ART: Aminosäure
- (C) STRANGFORM: Einzelstrang
- (D) TOPOLOGIE: linear

### (ii) ART DES MOLEKÜLS: Protein

### (vi) URSPRÜNGLICHE HERKUNFT:

- (A) ORGANISMUS: Pyrococcus horikoshii

### (xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 4:

Met His Asn Met Glu Glu Val Arg Glu Val Lys Val Leu Glu Lys Pro  
1 5 10 15

Trp Val Glu Lys Tyr Arg Pro Gln Arg Leu Asp Glu Ile Val Gly Gln  
20 25 30

# DE 198 40 771 A 1

Glu His Ile Val Lys Arg Leu Lys His Tyr Val Lys Thr Gly Ser Met  
 35 40 45  
 5 Pro His Leu Leu Phe Ala Gly Pro Pro Gly Val Gly Lys Cys Leu Thr  
 50 55 60  
 10 Gly Asp Thr Lys Val Ile Ala Asn Gly Gln Leu Phe Glu Leu Arg Glu  
 65 70 75 80  
 15 Leu Val Glu Lys Ile Ser Gly Gly Lys Phe Gly Pro Thr Pro Val Lys  
 85 90 95  
 20 Gly Leu Lys Val Ile Gly Ile Asp Glu Asp Gly Lys Leu Arg Glu Phe  
 100 105 110  
 25 Glu Val Gln Tyr Val Tyr Lys Asp Lys Thr Glu Arg Leu Ile Arg Ile  
 115 120 125  
 30 Arg Thr Arg Leu Gly Arg Glu Leu Lys Val Thr Pro Tyr His Pro Leu  
 130 135 140  
 35 Leu Val Asn Arg Arg Asn Gly Glu Ile Lys Trp Val Lys Ala Glu Glu  
 145 150 155 160  
 40 Leu Lys Pro Gly Asp Lys Leu Ala Val Pro Arg Phe Leu Pro Ile Val  
 165 170 175  
 45 Thr Gly Glu Asp Pro Leu Ala Glu Trp Leu Gly Tyr Phe Leu Gly Gly  
 180 185 190  
 50 Gly Tyr Ala Asp Ser Lys Glu Asn Leu Ile Met Phe Thr Asn Glu Asp  
 195 200 205  
 55 Pro Leu Leu Arg Gln Arg Phe Met Glu Leu Thr Glu Lys Leu Phe Ser  
 210 215 220  
 60 Asp Ala Arg Ile Arg Glu Ile Thr His Glu Asn Gly Thr Ser Lys Val  
 225 230 235 240  
 65 Tyr Val Asn Ser Lys Lys Ala Leu Lys Leu Val Asn Ser Leu Gly Asn  
 245 250 255  
 70 Ala His Ile Pro Lys Glu Cys Trp Arg Gly Ile Arg Ser Phe Leu Arg  
 260 265 270  
 75 Ala Tyr Phe Asp Cys Asn Gly Gly Val Lys Gly Asn Ala Ile Val Leu  
 275 280 285

# DE 198 40 771 A 1

Ala Thr Ala Ser Lys Glu Met Ser Gln Glu Ile Ala Tyr Ala Leu Ala  
290 295 300

Gly Phe Gly Ile Ile Ser Arg Ile Gln Glu Tyr Arg Val Ile Ile Ser  
305 310 315 320

Gly Ser Asp Asn Val Lys Lys Phe Leu Asn Glu Ile Gly Phe Ile Asn  
325 330 335

Arg Asn Lys Leu Glu Lys Ala Leu Lys Leu Val Lys Lys Asp Asp Pro  
340 345 350

Gly His Asp Gly Leu Glu Ile Asn Tyr Glu Leu Ile Ser Tyr Val Lys  
355 360 365

Asp Arg Leu Arg Leu Ser Phe Phe Asn Asp Lys Arg Ser Trp Ser Tyr  
370 375 380

Arg Glu Ala Lys Glu Ile Ser Trp Glu Leu Met Lys Glu Ile Tyr Tyr  
385 390 395 400

Arg Leu Asp Glu Leu Glu Lys Leu Lys Glu Ser Leu Ser Arg Gly Ile  
405 410 415

Leu Ile Asp Trp Asn Glu Val Ala Lys Arg Ile Glu Glu Val Ala Glu  
420 425 430

Glu Thr Gly Ile Arg Ala Asp Glu Leu Leu Glu Tyr Ile Glu Gly Lys  
435 440 445

Arg Lys Leu Ser Phe Lys Asp Tyr Ile Lys Ile Ala Lys Val Leu Gly  
450 455 460

Ile Asp Val Glu His Thr Ile Glu Ala Met Arg Val Phe Ala Arg Lys  
465 470 475 480

Tyr Ser Ser Tyr Ala Glu Ile Gly Arg Arg Leu Gly Thr Trp Asn Ser  
485 490 495

Ser Val Lys Thr Ile Leu Glu Ser Asn Ala Val Asn Val Glu Ile Leu  
500 505 510

Glu Arg Ile Arg Lys Ile Glu Leu Glu Leu Ile Glu Glu Ile Leu Ser  
515 520 525

Asp Glu Lys Leu Lys Glu Gly Ile Ala Tyr Leu Ile Phe Leu Ser Gln  
530 535 540

# DE 198 40 771 A 1

Asn Glu Leu Tyr Trp Asp Glu Ile Thr Lys Val Glu Glu Leu Arg Gly  
 545 550 555 560

5 Glu Phe Ile Ile Tyr Asp Leu His Val Pro Gly Tyr His Asn Phe Ile  
 565 570 575

10 Ala Gly Asn Met Pro Thr Val Val His Asn Thr Thr Ala Ala Leu Ala  
 580 585 590

15 Leu Ser Arg Glu Leu Phe Gly Glu Asn Trp Arg His Asn Phe Leu Glu  
 595 600 605

20 Leu Asn Ala Ser Asp Glu Arg Gly Ile Asn Val Ile Arg Glu Lys Val  
 610 615 620

25 Lys Glu Phe Ala Arg Thr Lys Pro Ile Gly Gly Ala Ser Phe Lys Ile  
 625 630 635 640

30 Ile Phe Leu Asp Glu Ala Asp Ala Leu Thr Gln Asp Ala Gln Gln Ala  
 645 650 655

35 Leu Arg Arg Thr Met Glu Met Phe Ser Ser Asn Val Arg Phe Ile Leu  
 660 665 670

40 Ser Cys Asn Tyr Ser Ser Lys Ile Ile Glu Pro Ile Gln Ser Arg Cys  
 675 680 685

45 Ala Ile Phe Arg Phe Arg Pro Leu Arg Asp Glu Asp Ile Ala Lys Arg  
 690 695 700

50 Leu Arg Tyr Ile Ala Glu Asn Glu Gly Leu Glu Leu Thr Glu Glu Gly  
 705 710 715 720

55 Leu Gln Ala Ile Leu Tyr Ile Ala Glu Gly Asp Met Arg Arg Ala Ile  
 725 730 735

60 Asn Ile Leu Gln Ala Ala Ala Ala Leu Asp Lys Lys Ile Thr Asp Glu  
 740 745 750

65 Asn Val Phe Met Val Ala Ser Arg Ala Arg Pro Glu Asp Ile Arg Glu  
 755 760 765

70 Met Met Leu Leu Ala Leu Lys Gly Asn Phe Leu Lys Ala Arg Glu Lys  
 770 775 780

75 Leu Arg Glu Ile Leu Leu Lys Gln Gly Leu Ser Gly Glu Asp Val Leu  
 785 790 795 800

Ile Gln Met His Lys Glu Val Phe Asn Leu Pro Ile Asp Glu Pro Thr  
805 810 815

Lys Val Tyr Leu Ala Asp Lys Ile Gly Glu Tyr Asn Phe Arg Leu Val  
820 825 830

Glu Gly Ala Asn Glu Met Ile Gln Leu Glu Ala Leu Leu Ala Gln Phe  
835 840 845

Thr Leu Val Gly Lys Lys Lys  
850 855

(2) ANGABEN ZU SEQ ID NO: 5:

(i) SEQUENZKENNZEICHEN:

(A) LÄNGE: 321 Aminosäuren

(B) ART: Aminosäure

(C) STRANGFORM: Einzelstrang

(D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: Protein

(vi) URSPRÜNGLICHE HERKUNFT:

(A) ORGANISMUS: Methanobacterium thermoautotrophicum

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 5:

Met Ile Ile Met Asn Gly Pro Trp Val Glu Lys Tyr Arg Pro Gln Lys  
1 5 10 15

Leu Asp Asp Ile Val Gly Gln Glu His Ile Ile Pro Arg Leu Lys Arg  
20 25 30

Tyr Val Glu Glu Lys Ser Met Pro Asn Leu Met Phe Thr Gly Pro Ala  
35 40 45

Gly Val Gly Lys Thr Thr Ala Ala Leu Ala Leu Ala Arg Glu Ile Leu  
50 55 60

Gly Glu Tyr Trp Arg Gln Asn Phe Leu Glu Leu Asn Ala Ser Asp Ala  
65 70 75 80

Arg Gly Ile Asp Thr Val Arg Thr Ser Ile Lys Asn Phe Cys Arg Leu  
85 90 95

# DE 198 40 771 A 1

Lys Pro Val Gly Ala Pro Phe Arg Ile Ile Phe Leu Asp Glu Val Asp  
 100 105 110

5 Asn Met Thr Lys Asp Ala Gln His Ala Leu Arg Arg Glu Met Glu Met  
 115 120 125

10 Tyr Thr Lys Thr Ser Ser Phe Ile Leu Ser Cys Asn Tyr Ser Ser Lys  
 130 135 140

Ile Ile Asp Pro Ile Gln Ser Arg Cys Ala Ile Phe Arg Phe Leu Pro  
 15 145 150 155 160

Leu Lys Gly His Gln Ile Ile Lys Arg Leu Glu Tyr Ile Ala Glu Lys  
 20 165 170 175

Glu Asn Leu Glu Tyr Glu Ala His Ala Leu Glu Thr Ile Val Tyr Phe  
 180 185 190

25 Ala Glu Gly Asp Leu Arg Lys Ala Ile Asn Leu Leu Gln Ser Ala Ala  
 195 200 205

30 Ser Leu Gly Glu Lys Ile Thr Glu Ser Ser Ile Tyr Asp Val Val Ser  
 210 215 220

Arg Ala Arg Pro Lys Asp Val Arg Lys Met Ile Lys Thr Ile Leu Asp  
 35 225 230 235 240

Gly Lys Phe Met Glu Ala Arg Asp Met Leu Arg Glu Ile Met Val Leu  
 245 250 255

40 Gln Gly Ile Ser Gly Glu Asp Met Val Thr Gln Ile Tyr Gln Glu Leu  
 260 265 270

45 Ser Arg Leu Ala Met Glu Gly Glu Val Asp Gly Asp Arg Tyr Val Gly  
 275 280 285

Leu Ile Asp Ala Ile Gly Glu Tyr Asp Phe Arg Ile Arg Glu Gly Ala  
 50 290 295 300

Asn Pro Arg Ile Gln Leu Glu Ala Leu Leu Ala Arg Phe Leu Glu His  
 55 305 310 315 320

Ala



## (2) ANGABEN ZU SEQ ID NO: 6:

## (i) SEQUENZKENNZEICHEN:

- (A) LÄNGE: 1148 Aminosäuren  
 (B) ART: Aminosäure  
 (C) STRANGFORM: Einzelstrang  
 (D) TOPOLOGIE: linear

## (ii) ART DES MOLEKÜLS: Protein

## (vi) URSPRÜNGLICHE HERKUNFT:

- (A) ORGANISMUS: Homo sapiens

## (xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 6:

Met Asp Ile Arg Lys Phe Phe Gly Val Ile Pro Ser Gly Lys Lys Leu  
 1 5 10 15

Val Ser Glu Thr Val Lys Lys Asn Glu Lys Thr Lys Ser Asp Glu Glu  
 20 25 30

Thr Leu Lys Ala Lys Lys Gly Ile Lys Glu Ile Lys Val Asn Ser Ser  
 35 40 45

Arg Lys Glu Asp Asp Phe Lys Gln Lys Gln Pro Ser Lys Lys Lys Arg  
 50 55 60

Ile Ile Tyr Asp Ser Asp Ser Glu Ser Glu Glu Thr Leu Gln Val Lys  
 65 70 75 80

Asn Ala Lys Lys Pro Pro Glu Lys Leu Pro Val Ser Ser Lys Pro Gly  
 85 90 95

Lys Ile Ser Arg Gln Asp Pro Val Thr Tyr Ile Ser Glu Thr Asp Glu  
 100 105 110

Glu Asp Asp Phe Met Cys Lys Lys Ala Ala Ser Lys Ser Lys Glu Asn  
 115 120 125

Gly Arg Ser Thr Asn Ser His Leu Gly Thr Ser Asn Met Lys Lys Asn  
 130 135 140

Glu Glu Asn Thr Lys Thr Lys Asn Lys Pro Leu Ser Pro Ile Lys Leu  
 145 150 155 160

# DE 198 40 771 A 1

Thr Pro Thr Ser Val Leu Asp Tyr Phe Gly Thr Gly Ser Val Gln Arg  
 165 170 175  
 5 Ser Asn Lys Lys Met Val Ala Ser Lys Arg Lys Glu Leu Ser Gln Asn  
 180 185 190  
 10 Thr Asp Glu Ser Gly Leu Asn Asp Glu Ala Ile Ala Lys Gln Leu Gln  
 195 200 205  
 Leu Asp Glu Asp Ala Glu Leu Glu Arg Gln Leu His Glu Asp Glu Glu  
 15 210 215 220  
 Phe Ala Arg Thr Leu Ala Met Leu Asp Glu Glu Pro Lys Thr Lys Lys  
 20 225 230 235 240  
 Ala Arg Lys Asp Thr Glu Ala Gly Glu Thr Phe Ser Ser Val Gln Ala  
 25 245 250 255  
 Asn Leu Ser Lys Ala Glu Lys His Lys Tyr Pro His Lys Val Lys Thr  
 260 265 270  
 Ala Gln Val Ser Asp Glu Arg Lys Ser Tyr Ser Pro Arg Lys Gln Ser  
 30 275 280 285  
 Lys Tyr Glu Ser Ser Lys Glu Ser Gln Gln His Ser Lys Ser Ser Ala  
 35 290 295 300  
 Asp Lys Ile Gly Glu Val Ser Ser Pro Lys Ala Ser Ser Lys Leu Ala  
 305 310 315 320  
 40 Ile Met Lys Arg Lys Lys Glu Ser Ser Tyr Lys Glu Ile Glu Pro Val  
 325 330 335  
 Ala Ser Lys Arg Lys Glu Asn Ala Ile Lys Leu Lys Gly Glu Thr Lys  
 45 340 345 350  
 Thr Pro Lys Lys Thr Lys Ser Ser Pro Ala Lys Lys Glu Ser Val Ser  
 50 355 360 365  
 Pro Glu Asp Ser Glu Lys Lys Arg Thr Asn Tyr Gln Ala Tyr Arg Ser  
 55 370 375 380  
 Tyr Leu Asn Arg Glu Gly Pro Lys Ala Leu Gly Ser Lys Glu Ile Pro  
 385 390 395 400  
 60 Lys Gly Ala Glu Asn Cys Leu Glu Gly Leu Ile Phe Val Ile Thr Gly  
 405 410 415  
 65

# DE 198 40 771 A 1

Val Leu Glu Ser Ile Glu Arg Asp Glu Ala Lys Ser Leu Ile Glu Arg	
420 425 430	
Tyr Gly Gly Lys Val Thr Gly Asn Val Ser Lys Lys Thr Asn Tyr Leu	5
435 440 445	
Val Met Gly Arg Asp Ser Gly Gln Ser Lys Ser Asp Lys Ala Ala Ala	10
450 455 460	
Leu Gly Thr Lys Ile Ile Asp Glu Asp Gly Leu Leu Asn Leu Ile Arg	15
465 470 475 480	
Thr Met Pro Gly Lys Lys Ser Lys Tyr Glu Ile Ala Val Glu Thr Glu	20
485 490 495	
Met Lys Lys Glu Ser Lys Leu Glu Arg Thr Pro Gln Lys Asn Val Gln	25
500 505 510	
Gly Lys Arg Lys Ile Ser Pro Ser Lys Lys Glu Ser Glu Ser Lys Lys	30
515 520 525	
Ser Arg Pro Thr Ser Lys Arg Asp Ser Leu Ala Lys Thr Ile Lys Lys	35
530 535 540	
Glu Thr Asp Val Phe Trp Lys Ser Leu Asp Phe Lys Glu Gln Val Ala	40
545 550 555 560	
Glu Glu Thr Ser Gly Asp Ser Lys Ala Arg Asn Leu Ala Asp Asp Ser	45
565 570 575	
Ser Glu Asn Lys Val Glu Asn Leu Leu Trp Val Asp Lys Tyr Lys Pro	50
580 585 590	
Thr Ser Leu Lys Thr Ile Ile Gly Gln Gln Gly Asp Gln Ser Cys Ala	55
595 600 605	
Asn Lys Leu Leu Arg Trp Leu Arg Asn Trp Gln Lys Ser Ser Ser Glu	60
610 615 620	
Asp Lys Lys His Ala Ala Lys Phe Gly Lys Phe Ser Gly Lys Asp Asp	65
625 630 635 640	
Gly Ser Ser Phe Lys Ala Ala Leu Leu Ser Gly Pro Pro Gly Val Gly	
645 650 655	
Lys Thr Thr Thr Ala Ser Leu Val Cys Gln Glu Leu Gly Tyr Ser Tyr	
660 665 670	

# DE 198 40 771 A 1

Val Glu Leu Asn Ala Ser Asp Thr Arg Ser Lys Ser Ser Leu Lys Ala  
675 680 685

5 Ile Val Ala Glu Ser Leu Asn Asn Thr Ser Ile Lys Gly Phe Tyr Ser  
690 695 700

10 Asn Gly Ala Ala Ser Ser Val Ser Thr Lys His Ala Leu Ile Met Asp  
705 710 715 720

Glu Val Asp Gly Met Ala Gly Asn Glu Asp Arg Gly Gly Ile Gln Glu  
15 725 730 735

Leu Ile Gly Leu Ile Lys His Thr Lys Ile Pro Ile Ile Cys Met Cys  
20 740 745 750

Asn Asp Arg Asn His Pro Lys Ile Arg Ser Leu Val His Tyr Cys Phe  
25 755 760 765

Asp Leu Arg Phe Gln Arg Pro Arg Val Glu Gln Ile Lys Gly Ala Met  
30 770 775 780

Met Ser Ile Ala Phe Lys Glu Gly Leu Lys Ile Pro Pro Pro Ala Met  
35 785 790 795 800

Asn Glu Ile Ile Leu Gly Ala Asn Gln Asp Ile Arg Gln Val Leu His  
40 805 810 815

Asn Leu Ser Met Trp Cys Ala Arg Ser Lys Ala Leu Thr Tyr Asp Gln  
45 820 825 830

Ala Lys Ala Asp Ser His Arg Ala Lys Lys Asp Ile Lys Met Gly Pro  
50 835 840 845

Phe Asp Val Ala Arg Lys Val Phe Ala Ala Gly Glu Glu Thr Ala His  
55 850 855 860

Met Ser Leu Val Asp Lys Ser Asp Leu Phe Phe His Asp Tyr Ser Ile  
60 865 870 875 880

Ala Pro Leu Phe Val Gln Glu Asn Tyr Ile His Val Lys Pro Val Ala  
65 885 890 895

Ala Gly Gly Asp Met Lys Lys His Leu Met Leu Leu Ser Arg Ala Ala  
900 905 910

Asp Ser Ile Cys Asp Gly Asp Leu Val Asp Ser Gln Ile Arg Ser Lys  
915 920 925

Gln Asn Trp Ser Leu Leu Pro Ala Gln Ala Ile Tyr Ala Ser Val Leu  
930 935 940

Pro Gly Glu Leu Met Arg Gly Tyr Met Thr Gln Phe Pro Thr Phe Pro  
945 950 955 960

Ser Trp Leu Gly Lys His Ser Ser Thr Gly Lys His Asp Arg Ile Val  
965 970 975

Gln Asp Leu Ala Leu His Met Ser Leu Arg Thr Tyr Ser Ser Lys Arg  
980 985 990

Thr Val Asn Met Asp Tyr Leu Ser Leu Leu Arg Asp Ala Leu Val Gln  
995 1000 1005

Pro Leu Thr Ser Gln Gly Val Asp Gly Val Gln Asp Val Val Ala Leu  
1010 1015 1020

Met Asp Thr Tyr Tyr Leu Met Lys Glu Asp Phe Glu Asn Ile Met Glu  
1025 1030 1035 1040

Ile Ser Ser Trp Gly Gly Lys Pro Ser Pro Phe Ser Lys Leu Asp Pro  
1045 1050 1055

Lys Val Lys Ala Ala Phe Thr Arg Ala Tyr Asn Lys Glu Ala His Leu  
1060 1065 1070

Thr Pro Tyr Ser Leu Gln Ala Ile Lys Ala Ser Arg His Ser Thr Ser  
1075 1080 1085

Pro Ser Leu Asp Ser Glu Tyr Asn Glu Glu Leu Asn Glu Asp Asp Ser  
1090 1095 1100

Gln Ser Asp Glu Lys Asp Gln Asp Ala Ile Glu Thr Asp Ala Met Ile  
1105 1110 1115 1120

Lys Lys Lys Thr Lys Ser Ser Lys Pro Ser Lys Pro Glu Lys Asp Lys  
1125 1130 1135

Glu Pro Arg Lys Gly Lys Gly Lys Ser Ser Lys Lys  
1140 1145

(2) ANGABEN ZU SEQ ID NO: 7:

(i) SEQUENZKENNZEICHEN:

(A) LÄNGE: 479 Aminosäuren

(B) ART: Aminosäure

(C) STRANGFORM: Einzelstrang

(D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: Protein

(vi) URSPRÜNGLICHE HERKUNFT:

(A) ORGANISMUS: Archaeoglobus fulgidus

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 7:

Met Leu Trp Val Glu Lys Tyr Arg Pro Lys Thr Leu Glu Glu Val Val  
1 5 10 15

Ala Asp Lys Ser Ile Ile Thr Arg Val Ile Lys Trp Ala Lys Ser Trp  
20 25 30

Lys Arg Gly Ser Lys Pro Leu Leu Leu Ala Gly Pro Pro Gly Val Gly  
35 40 45

Lys Thr Ser Leu Ala Leu Ala Leu Ala Asn Thr Met Gly Trp Glu Ala  
50 55 60

Val Glu Leu Asn Ala Ser Asp Gln Arg Ser Trp Arg Val Ile Glu Arg  
65 70 75 80

Ile Val Gly Glu Gly Ala Phe Asn Glu Thr Ile Ser Asp Glu Gly Glu  
85 90 95

Phe Leu Ser Ser Arg Ile Gly Lys Leu Lys Leu Ile Ile Leu Asp Glu  
100 105 110

Val Asp Asn Ile His Lys Lys Glu Asp Val Gly Gly Glu Ala Ala Leu  
115 120 125

Ile Arg Leu Ile Lys Arg Lys Pro Ala Gln Pro Leu Ile Leu Ile Ala  
130 135 140

Asn Asp Pro Tyr Lys Leu Ser Pro Glu Leu Arg Asn Leu Cys Glu Met  
145 150 155 160

Ile Asn Phe Lys Arg Leu Thr Lys Gln Gln Val Ala Arg Val Leu Glu  
165 170 175

Arg Ile Ala Leu Lys Glu Gly Ile Lys Val Asp Lys Ser Val Leu Leu  
180 185 190

# DE 198 40 771 A 1

Lys Ile Ala Glu Asn Ala Gly Gly Asp Leu Arg Ala Ala Ile Asn Asp  
195 200 205

Phe Gln Ala Leu Ala Glu Gly Lys Glu Glu Leu Lys Pro Glu Asp Val  
210 215 220

Phe Leu Thr Lys Arg Thr Gln Glu Lys Asp Ile Phe Arg Val Met Gln  
225 230 235 240

Met Ile Phe Lys Thr Lys Asn Pro Ala Val Tyr Asn Glu Ala Met Leu  
245 250 255

Leu Asp Glu Ser Pro Glu Asp Val Ile His Trp Val Asp Glu Asn Leu  
260 265 270

Pro Leu Glu Tyr Ser Gly Val Glu Leu Val Asn Ala Tyr Glu Ala Leu  
275 280 285

Ser Arg Ala Asp Ile Phe Leu Gly Arg Val Arg Arg Arg Gln Phe Tyr  
290 295 300

Arg Leu Trp Lys Tyr Ala Ser Tyr Leu Met Thr Val Gly Val Gln Gln  
305 310 315 320

Met Lys Glu Glu Pro Lys Lys Gly Phe Thr Arg Tyr Arg Arg Pro Ala  
325 330 335

Val Trp Gln Met Leu Phe Gln Leu Arg Gln Lys Arg Glu Met Thr Arg  
340 345 350

Lys Ile Leu Glu Lys Ile Gly Lys Tyr Ser His Leu Ser Met Arg Lys  
355 360 365

Ala Arg Thr Glu Met Phe Pro Val Ile Lys Leu Leu Leu Lys Glu Leu  
370 375 380

Asp Val Asp Lys Ala Ala Thr Ile Ala Ala Phe Tyr Glu Phe Thr Lys  
385 390 395 400

Glu Glu Leu Glu Phe Leu Val Gly Glu Lys Gly Asp Glu Ile Trp Lys  
405 410 415

Tyr Val Glu Lys His Gly Met His Arg Ile Glu Asp Glu Thr Phe Leu  
420 425 430

Glu Ser Phe Val Lys Ala Glu Lys Glu Glu Lys Glu Glu Ser Val Glu  
435 440 445

Glu Val Ala Glu Glu Lys Pro Glu Glu Glu Arg Glu Glu Pro Arg Ala  
450 455 460

Arg Lys Lys Ala Gly Lys Asn Leu Thr Leu Asp Ser Phe Phe Ser  
465 470 475

(2) ANGABEN ZU SEQ ID NO: 8:

(i) SEQUENZKENNZEICHEN:

- (A) LÄNGE: 516 Aminosäuren
- (B) ART: Aminosäure
- (C) STRANGFORM: Einzelstrang
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: Protein

(vi) URSPRÜNGLICHE HERKUNFT:

- (A) ORGANISMUS: Methanococcus jannaschii

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 8:

Met Leu Ser Trp Val Glu Lys Tyr Arg Pro Lys Ser Leu Lys Asp Val  
1 5 10 15

Ala Gly His Glu Lys Val Lys Glu Lys Leu Lys Thr Trp Ile Glu Ser  
20 25 30

Tyr Leu Lys Gly Glu Thr Pro Lys Pro Ile Leu Leu Val Gly Pro Pro  
35 40 45

Gly Cys Gly Lys Thr Thr Leu Ala Tyr Ala Leu Ala Asn Asp Tyr Gly  
50 55 60

Phe Glu Val Ile Glu Leu Asn Ala Ser Asp Lys Arg Asn Ser Ser Ala  
65 70 75 80

Ile Lys Lys Val Val Gly His Ala Ala Thr Ser Ser Ser Ile Phe Gly  
85 90 95

Lys Lys Phe Leu Ile Val Leu Asp Glu Val Asp Gly Ile Ser Gly Lys  
100 105 110

Glu Asp Ala Gly Gly Val Ser Glu Leu Ile Lys Val Ile Lys Lys Ala  
115 120 125



# DE 198 40 771 A 1

Lys Asn Pro Ile Ile Leu Thr Ala Asn Asp Ala Tyr Ala Pro Ser Ile  
130 135 140

Arg Ser Leu Leu Pro Tyr Val Glu Val Ile Gln Leu Asn Pro Val His  
145 150 155 160

Thr Asn Ser Val Tyr Lys Val Leu Lys Lys Ile Ala Glu Lys Glu Gly  
165 170 175

Leu Asp Val Asp Asp Lys Thr Leu Lys Met Ile Ala Gln His Ser Ala  
180 185 190

Gly Asp Leu Arg Ser Ala Ile Asn Asp Leu Glu Ala Leu Ala Leu Ser  
195 200 205

Gly Asp Leu Ser Tyr Glu Ala Ala Gln Lys Leu Pro Asp Arg Lys Arg  
210 215 220

Glu Ala Asn Ile Phe Asp Ala Leu Arg Val Ile Leu Lys Thr Thr His  
225 230 235 240

Tyr Gly Ile Ala Thr Thr Ala Leu Met Asn Val Asp Glu Thr Pro Asp  
245 250 255

Val Val Ile Glu Trp Ile Ala Glu Asn Val Pro Lys Glu Tyr Glu Lys  
260 265 270

Pro Glu Glu Val Ala Arg Ala Phe Glu Tyr Leu Ser Lys Ala Asp Arg  
275 280 285

Tyr Leu Gly Arg Val Met Arg Arg Gln Asn Tyr Ser Phe Trp Lys Tyr  
290 295 300

Ala Thr Thr Leu Met Thr Ala Gly Val Ala Leu Ser Lys Asp Glu Lys  
305 310 315 320

Tyr Arg Lys Trp Thr Pro Tyr Ser Tyr Pro Lys Ile Phe Arg Leu Leu  
325 330 335

Thr Lys Thr Lys Ala Glu Arg Glu Ile Leu Asn Lys Ile Leu Lys Lys  
340 345 350

Ile Gly Glu Lys Thr His Thr Ser Ser Lys Arg Ala Arg Phe Asp Leu  
355 360 365

Gln Met Leu Lys Leu Leu Ala Lys Glu Asn Pro Ser Val Ala Ala Asp  
370 375 380

# DE 198 40 771 A 1

Leu Val Asp Tyr Phe Glu Ile Lys Glu Asp Glu Leu Lys Val Leu Val  
385 390 395 400

Gly Asp Lys Leu Ala Ser Glu Ile Leu Lys Ile Leu Lys Glu Lys Lys  
405 410 415

Lys Leu Glu Arg Lys Lys Lys Lys Glu Lys Glu Lys Leu Glu Lys Glu  
420 425 430

Lys Lys Lys Glu Glu Lys Ala Lys Glu Lys Gln Ser Asn Leu Ile Ile  
435 440 445

Gln Pro Lys Glu Ile Lys Glu Glu Val Lys Ala Glu Val Glu Lys Lys  
450 455 460

Glu Glu Val Lys Glu Lys Ile Val Glu Lys Pro Lys Ala Glu Glu Val  
465 470 475 480

Lys Glu Lys Ser Lys Thr Glu Glu Lys Glu Thr Lys Lys Asp Lys Lys  
485 490 495

Lys Gly Lys Lys Lys Lys Glu Asp Lys Gly Lys Gln Leu Thr Leu Asp  
500 505 510

Ala Phe Phe Lys  
515

## (2) ANGABEN ZU SEQ ID NO: 9:

### (i) SEQUENZKENNZEICHEN:

- (A) LÄNGE: 468 Aminosäuren
- (B) ART: Aminosäure
- (C) STRANGFORM: Einzelstrang
- (D) TOPOLOGIE: linear

### (ii) ART DES MOLEKÜLS: Protein

### (vi) URSPRÜNGLICHE HERKUNFT:

- (A) ORGANISMUS: Pyrococcus horikoshii

### (xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 9:

Met Pro Asp Val Pro Trp Ile Glu Lys Tyr Arg Pro Arg Lys Leu Ser  
1 5 10 15

# DE 198 40 771 A 1

Glu Ile Val Asn Gln Glu Gln Ala Leu Glu Lys Val Arg Ala Trp Ile	
20 25 30	
Glu Ser Trp Leu His Gly Asn Pro Pro Lys Lys Lys Ala Leu Leu Leu	5
35 40 45	
Ala Gly Pro Pro Gly Ser Gly Lys Thr Thr Thr Val Tyr Ala Leu Ala	10
50 55 60	
His Glu Tyr Asn Phe Glu Val Ile Glu Leu Asn Ala Ser Asp Glu Arg	15
65 70 75 80	
Thr Tyr Asn Lys Ile Ala Arg Tyr Val Gln Ala Ala Tyr Thr Met Asp	20
85 90 95	
Ile Met Gly Lys Arg Arg Lys Ile Ile Phe Leu Asp Glu Ala Asp Asn	25
100 105 110	
Ile Glu Pro Ser Gly Ala Pro Glu Ile Ala Lys Leu Ile Asp Lys Ala	30
115 120 125	
Arg Asn Pro Ile Ile Met Ala Ala Asn His Tyr Trp Glu Val Pro Lys	35
130 135 140	
Glu Ile Arg Asp Arg Ala Glu Leu Val Glu Tyr Lys Arg Leu Asn Gln	40
145 150 155 160	
Arg Asp Val Ile Ser Ala Leu Val Arg Ile Leu Lys Arg Glu Gly Ile	45
165 170 175	
Thr Val Pro Lys Glu Ile Leu Thr Glu Ile Ala Lys Arg Ser Ser Gly	50
180 185 190	
Asp Leu Arg Ala Ala Ile Asn Asp Leu Gln Thr Ile Val Ala Gly Gly	55
195 200 205	
Tyr Glu Asp Ala Lys Tyr Val Leu Ala Tyr Arg Asp Val Glu Lys Thr	60
210 215 220	
Val Phe Gln Ser Leu Gly Met Val Phe Ser Ser Asp Asn Ala Lys Arg	65
225 230 235 240	
Ala Lys Leu Ala Leu Met Asn Leu Asp Met Ser Pro Asp Glu Phe Leu	
245 250 255	
Leu Trp Val Asp Glu Asn Ile Pro His Met Tyr Leu Lys Pro Glu Glu	
260 265 270	

# DE 198 40 771 A 1

Met Ala Arg Ala Tyr Glu Ala Ile Ser Arg Ala Asp Ile Tyr Leu Gly  
275 280 285

Arg Ala Gln Arg Thr Gly Asn Tyr Ser Leu Trp Lys Tyr Ala Ile Asp  
290 295 300

Met Met Thr Ala Gly Val Ala Val Ala Gly Thr Lys Lys Lys Gly Phe  
305 310 315 320

Ala Lys Phe Tyr Pro Pro Asn Thr Leu Lys Met Leu Ala Glu Ser Lys  
325 330 335

Glu Glu Arg Ser Ile Arg Asp Ser Ile Ile Lys Lys Ile Met Lys Glu  
340 345 350

Met His Met Ser Lys Leu Glu Ala Leu Glu Thr Met Lys Ile Leu Arg  
355 360 365

Thr Ile Phe Glu Asn Asn Leu Asp Leu Ala Ala His Phe Thr Val Phe  
370 375 380

Leu Glu Leu Thr Glu Lys Glu Val Glu Phe Leu Ala Gly Lys Glu Lys  
385 390 395 400

Ala Gly Thr Ile Trp Gly Lys Thr Leu Ser Ile Arg Arg Arg Ile Lys  
405 410 415

Glu Thr Glu Lys Ile Glu Glu Lys Ala Val Glu Glu Lys Val Glu Glu  
420 425 430

Glu Glu Ala Glu Glu Glu Glu Glu Glu Arg Lys Glu Glu Glu Lys  
435 440 445

Pro Lys Ala Glu Lys Lys Lys Gly Lys Gln Val Thr Leu Phe Asp Phe  
450 455 460

Ile Lys Lys Asn  
465

## (2) ANGABEN ZU SEQ ID NO: 10:

### (i) SEQUENZKENNZEICHEN:

- (A) LÄNGE: 479 Aminosäuren
- (B) ART: Aminosäure
- (C) STRANGFORM: Einzelstrang
- (D) TOPOLOGIE: linear

### (ii) ART DES MOLEKÜLS: Protein

## (vi) URSPRÜNGLICHE HERKUNFT:

(A) ORGANISMUS: *Methanobacterium thermoautotrophicum*

## (xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 10:

Met Ser Trp Thr Glu Lys Tyr Arg Pro Gly Ser Phe Asp Glu Val Val  
1 5 10 15

Gly Asn Gln Lys Val Ile Ala Glu Ile Lys Glu Trp Ile Lys Ala Trp  
20 25 30

Lys Ala Gly Lys Pro Gln Lys Pro Leu Leu Val Gly Pro Pro Gly  
35 40 45

Thr Gly Lys Thr Thr Leu Ala His Ile Ile Gly Lys Glu Phe Ser Asp  
50 55 60

Thr Leu Glu Leu Asn Ala Ser Asp Arg Arg Ser Gln Asp Ala Leu Met  
65 70 75 80

Arg Ser Ala Gly Glu Ala Ser Ala Thr Arg Ser Leu Phe Asn His Asp  
85 90 95

Leu Lys Leu Ile Ile Leu Asp Glu Val Asp Gly Ile His Gly Asn Glu  
100 105 110

Asp Arg Gly Gly Val Gln Ala Ile Asn Arg Ile Ile Lys Glu Ser Arg  
115 120 125

His Pro Met Val Leu Thr Ala Asn Asp Pro Tyr Ser Lys Arg Leu Gln  
130 135 140

Ser Ile Lys Pro Arg Cys Arg Val Leu Asn Leu Arg Lys Val His Thr  
145 150 155 160

Ser Ser Ile Ala Ala Ala Leu Arg Arg Ile Cys Arg Ala Glu Gly Ile  
165 170 175

Glu Cys Pro Asp Asp Val Leu Arg Glu Leu Ala Lys Arg Ser Arg Gly  
180 185 190

Asp Leu Arg Ser Ala Ile Asn Asp Leu Glu Ala Met Ala Glu Gly Glu  
195 200 205

Glu Arg Ile Gly Glu Glu Leu Leu Lys Met Gly Glu Lys Asp Ala Thr  
210 215 220

# DE 198 40 771 A 1

Ser Asn Leu Phe Asp Ala Val Arg Ala Val Leu Lys Ser Arg Asp Val  
225 230 235 240

5 Ser Lys Val Arg Glu Ala Met Arg Val Asp Asp Asp Pro Thr Leu Val  
245 250 255

10 Leu Glu Phe Ile Ala Glu Asn Val Pro Arg Glu Tyr Glu Lys Pro Asn  
260 265 270

15 Glu Ile Ser Arg Ala Tyr Asp Met Leu Ser Arg Ala Asp Ile Phe Phe  
275 280 285

20 Gly Arg Ala Val Arg Thr Arg Asn Tyr Thr Tyr Trp Arg Tyr Ala Ser  
290 295 300

Glu Leu Met Gly Pro Gly Val Ala Leu Ala Lys Asp Lys Thr Tyr Arg  
305 310 315 320

25 Lys Phe Val Arg Tyr Thr Gly Ser Ser Ser Phe Arg Ile Leu Gly Lys  
325 330 335

30 Thr Arg Lys Gln Arg Ser Leu Arg Asp Ser Val Ala Ala Lys Met Ala  
340 345 350

35 Gly Lys Met His Ile Ser Pro Lys Val Ala Ile Ser Met Phe Pro Tyr  
355 360 365

Met Glu Ile Leu Phe Glu Asn Asp Glu Met Ala Tyr Asp Ile Ser Glu  
370 375 380

40 Phe Leu Glu Leu Arg Asp Glu Glu Ile Lys Leu Phe Arg Lys Arg Lys  
385 390 395 400

45 Ile Lys Ala Pro Lys Arg Lys Lys Thr Pro Arg Lys Ala Glu Ile Lys  
405 410 415

50 Val Gly Pro Leu Tyr Ser Gln Lys Lys Asp Lys Gly Ala Asp Lys Ser  
420 425 430

55 Ile Asn Asp Lys Ala Thr Asp Lys Ser Ala Lys Thr Pro Ile Lys Ser  
435 440 445

Ser Lys Lys Asp Asp Arg Pro Arg Asp Glu Ser Ser Ser Ser Ser Asp  
450 455 460

60 Asp Lys Lys Pro Lys Glu Lys Gln Thr Ser Leu Phe Gln Phe Ser  
465 470 475

65

## (2) ANGABEN ZU SEQ ID NO: 11:

## (i) SEQUENZKENNZEICHEN:

(A) LÄNGE: 261 Aminosäuren 5

(B) ART: Aminosäure

(C) STRANGFORM: Einzelstrang

(D) TOPOLOGIE: linear 10

## (ii) ART DES MOLEKÜLS: Protein

## (vi) URSPRÜNGLICHE HERKUNFT: 15

(A) ORGANISMUS: Homo sapiens

## (xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 11: 20

Met Phe Glu Ala Arg Leu Val Gln Gly Ser Ile Leu Lys Lys Val Leu  
 1 5 10 15 25

Glu Ala Leu Lys Asp Leu Ile Asn Glu Ala Cys Trp Asp Ile Ser Ser  
 20 25 30 30

Ser Gly Val Asn Leu Gln Ser Met Asp Ser Ser His Val Ser Leu Val  
 35 40 45

Gln Leu Thr Leu Arg Ser Glu Gly Phe Asp Thr Tyr Arg Cys Asp Arg  
 50 55 60 35

Asn Leu Ala Met Gly Val Asn Leu Thr Ser Met Ser Lys Ile Leu Lys  
 65 70 75 80 40

Cys Ala Gly Asn Glu Asp Ile Ile Thr Leu Arg Ala Glu Asp Asn Ala  
 85 90 95 45

Asp Thr Leu Ala Leu Val Phe Glu Ala Pro Asn Gln Glu Lys Val Ser  
 100 105 110 50

Asp Tyr Glu Met Lys Leu Met Asp Leu Asp Val Glu Gln Leu Gly Ile  
 115 120 125 50

Pro Glu Gln Glu Tyr Ser Cys Val Val Lys Met Pro Ser Gly Glu Phe  
 130 135 140 55

Ala Arg Ile Cys Arg Asp Leu Ser His Ile Gly Asp Ala Val Val Ile  
 145 150 155 160 60

65

# DE 198 40 771 A 1

Ser Cys Ala Lys Asp Gly Val Lys Phe Ser Ala Ser Gly Glu Leu Gly  
165 170 175

Asn Gly Asn Ile Lys Leu Ser Gln Thr Ser Asn Val Asp Lys Glu Glu  
180 185 190

Glu Ala Val Thr Ile Glu Met Asn Glu Pro Val Gln Leu Thr Phe Ala  
195 200 205

Leu Arg Tyr Leu Asn Phe Phe Thr Lys Ala Thr Pro Leu Ser Ser Thr  
210 215 220

Val Thr Leu Ser Met Ser Ala Asp Val Pro Leu Val Val Glu Tyr Lys  
225 230 235 240

Ile Ala Asp Met Gly His Leu Lys Tyr Tyr Leu Ala Pro Lys Ile Glu  
245 250 255

Asp Glu Glu Gly Ser  
260

## (2) ANGABEN ZU SEQ ID NO: 12:

### (i) SEQUENZKENNZEICHEN:

(A) LÄNGE: 245 Aminosäuren

(B) ART: Aminosäure

(C) STRANGFORM: Einzelstrang

(D) TOPOLOGIE: linear

### (ii) ART DES MOLEKÜLS: Protein

### (vi) URSPRÜNGLICHE HERKUNFT:

(A) ORGANISMUS: Archaeoglobus fulgidus

### (xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 12:

Met Ile Asp Val Ile Met Thr Gly Glu Leu Leu Lys Thr Val Thr Arg  
1 5 10 15

Ala Ile Val Ala Leu Val Ser Glu Ala Arg Ile His Phe Leu Glu Lys  
20 25 30

Gly Leu His Ser Arg Ala Val Asp Pro Ala Asn Val Ala Met Val Ile  
35 40 45



# DE 198 40 771 A 1

Val Asp Ile Pro Lys Asp Ser Phe Glu Val Tyr Asn Ile Asp Glu Glu

50 55 60

Lys Thr Ile Gly Val Asp Met Asp Arg Ile Phe Asp Ile Ser Lys Ser

65 70 75 80

Ile Ser Thr Lys Asp Leu Val Glu Leu Ile Val Glu Asp Glu Ser Thr

85 90 95

Leu Lys Val Lys Phe Gly Ser Val Glu Tyr Lys Val Ala Leu Ile Asp

100 105 110

Pro Ser Ala Ile Arg Lys Glu Pro Arg Ile Pro Glu Leu Glu Leu Pro

115 120 125

Ala Lys Ile Val Met Asp Ala Gly Glu Phe Lys Lys Ala Ile Ala Ala

130 135 140

Ala Asp Lys Ile Ser Asp Gln Val Ile Phe Arg Ser Asp Lys Glu Gly

145 150 155 160

Phe Arg Ile Glu Ala Lys Gly Asp Val Asp Ser Ile Val Phe His Met

165 170 175

Thr Glu Thr Glu Leu Ile Glu Phe Asn Gly Gly Glu Ala Arg Ser Met

180 185 190

Phe Ser Val Asp Tyr Leu Lys Glu Phe Cys Lys Val Ala Gly Ser Gly

195 200 205

Asp Leu Leu Thr Ile His Leu Gly Thr Asn Tyr Pro Val Arg Leu Val

210 215 220

Phe Glu Leu Val Gly Gly Arg Ala Lys Val Glu Tyr Ile Leu Ala Pro

225 230 235 240

Arg Ile Glu Ser Glu

245

## (2) ANGABEN ZU SEQ ID NO: 13:

### (i) SEQUENZKENNZEICHEN:

(A) LÄNGE: 247 Aminosäuren

(B) ART: Aminosäure

(C) STRANGFORM: Einzelstrang

(D) TOPOLOGIE: linear

### (ii) ART DES MOLEKÜLS: Protein

## (vi) URSPRÜNGLICHE HERKUNFT:

(A) ORGANISMUS: *Methanococcus jannaschii*

5

## (xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 13:

10

Met Phe Arg Gly Val Met Glu Ser Ala Lys Glu Phe Lys Lys Val Val  
 1 5 10 15

15

Asp Thr Ile Ser Thr Leu Leu Asp Glu Ile Cys Phe Glu Val Asp Glu  
 20 25 30

20

Glu Gly Ile Lys Ala Ser Ala Met Asp Pro Ser His Val Ala Leu Val  
 35 40 45

25

Ser Leu Glu Ile Pro Arg Leu Ala Phe Glu Glu Tyr Glu Ala Asp Ser  
 50 55 60

His Asp Ile Gly Ile Asp Leu Glu Ala Phe Lys Lys Val Met Asn Arg  
 65 70 75 80

30

Ala Lys Ala Lys Asp Arg Leu Ile Leu Glu Leu Asp Glu Glu Lys Asn  
 85 90 95

35

Lys Leu Asn Val Ile Phe Glu Asn Thr Gly Lys Arg Lys Phe Ser Leu  
 100 105 110

40

Ala Leu Leu Asp Ile Ser Ala Ser Ser Val Lys Val Pro Glu Ile Glu  
 115 120 125

45

Tyr Pro Asn Val Ile Met Ile Lys Gly Asp Ala Phe Lys Glu Ala Leu  
 130 135 140

Lys Asp Ala Asp Leu Phe Ser Asp Tyr Val Ile Leu Lys Ala Asp Glu  
 145 150 155 160

50

Asp Lys Phe Val Ile His Ala Lys Gly Asp Leu Asn Glu Asn Glu Ala  
 165 170 175

55

Ile Phe Glu Lys Asp Ser Ser Ala Ile Ile Ser Leu Glu Val Lys Glu  
 180 185 190

60

Glu Ala Lys Ser Ala Phe Asn Leu Asp Tyr Leu Met Asp Met Val Lys  
 195 200 205

Gly Val Ser Ser Gly Asp Ile Ile Lys Ile Tyr Leu Gly Asn Asp Met  
 210 215 220

65

Pro Leu Lys Leu Glu Tyr Ser Ile Ala Gly Val Asn Leu Thr Phe Leu  
 225                      230                      235                      240

Leu Ala Pro Arg Ile Glu Gly  
 245

(2) ANGABEN ZU SEQ ID NO: 14:

(i) SEQUENZKENNZEICHEN:

(A) LÄNGE: 249 Aminosäuren

(B) ART: Aminosäure

(C) STRANGFORM: Einzelstrang

(D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: Protein

(vi) URSPRÜNGLICHE HERKUNFT:

(A) ORGANISMUS: Pyrococcus horikoshii

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 14:

Met Pro Phe Glu Ile Val Phe Glu Gly Ala Lys Glu Phe Ala Gln Leu  
 1                      5                      10                      15

Ile Glu Thr Ala Ser Arg Leu Ile Asp Glu Ala Ala Phe Lys Val Thr  
 20                      25                      30

Glu Glu Gly Ile Ser Met Arg Ala Met Asp Pro Ser Arg Val Val Leu  
 35                      40                      45

Ile Asp Leu Asn Leu Pro Ser Ser Ile Phe Ser Lys Tyr Glu Val Asp  
 50                      55                      60

Gly Glu Glu Thr Ile Gly Val Asn Met Asp His Leu Lys Lys Val Leu  
 65                      70                      75                      80

Lys Arg Gly Lys Ala Lys Asp Thr Leu Ile Leu Arg Lys Gly Glu Glu  
 85                      90                      95

Asn Phe Leu Glu Ile Ser Leu Gln Gly Thr Ala Thr Arg Thr Phe Arg  
 100                      105                      110

Leu Pro Leu Ile Asp Val Glu Glu Ile Glu Val Glu Leu Pro Asp Leu  
 115                      120                      125

DE 198 40 771 A 1

Pro Tyr Thr Ala Lys Val Val Val Leu Gly Glu Val Leu Lys Glu Ala  
130 135 140

Val Lys Asp Ala Ser Leu Val Ser Asp Ser Ile Lys Phe Met Ala Lys  
145 150 155 160

Glu Asn Glu Phe Ile Met Arg Ala Glu Gly Glu Thr Gln Glu Val Glu  
165 170 175

Val Lys Leu Thr Leu Glu Asp Glu Gly Leu Leu Asp Ile Glu Val Gln  
180 185 190

Glu Glu Thr Lys Ser Ala Tyr Gly Val Ser Tyr Leu Ala Asp Met Val  
195 200 205

Lys Gly Ile Gly Lys Ala Asp Glu Val Thr Met Arg Phe Gly Asn Glu  
210 215 220

Met Pro Met Gln Met Glu Tyr Tyr Ile Arg Asp Glu Gly Arg Leu Thr  
225 230 235 240

Phe Leu Leu Ala Pro Arg Val Glu Glu  
245

(2) ANGABEN ZU SEQ ID NO: 15:

(i) SEQUENZKENNZEICHEN:

(A) LÄNGE: 244 Aminosäuren

(B) ART: Aminosäure

(C) STRANGFORM: Einzelstrang

(D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: Protein

(vi) URSPRÜNGLICHE HERKUNFT:

(A) ORGANISMUS: Methanobacterium thermoautotrophicum

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 15:

Met Phe Lys Ala Glu Leu Asn Asp Pro Asn Ile Leu Arg Thr Ser Phe  
1 5 10 15

Asp Ala Ile Ser Ser Ile Val Asp Glu Val Gln Ile Gln Leu Ser Ala  
20 25 30

# DE 198 40 771 A 1

Glu Gly Leu Arg Leu Asp Ala Leu Asp Arg Ser His Ile Thr Tyr Val  
35 40 45

His Leu Glu Leu Lys Ala Glu Leu Phe Asp Glu Tyr Val Cys Asp Glu  
50 55 60

Pro Glu Arg Ile Asn Val Asp Thr Glu Glu Leu Met Lys Val Leu Lys  
65 70 75 80

Arg Ala Lys Ala Asn Asp Arg Val Ile Leu Ser Thr Asp Glu Gly Asn  
85 90 95

Leu Ile Ile Gln Phe Glu Gly Glu Ala Val Arg Thr Phe Lys Ile Arg  
100 105 110

Leu Ile Asp Ile Glu Tyr Glu Thr Pro Ser Pro Pro Glu Ile Glu Tyr  
115 120 125

Glu Asn Glu Phe Glu Val Pro Phe Gln Leu Leu Lys Asp Ser Ile Ala  
130 135 140

Asp Ile Asp Ile Phe Ser Asp Lys Ile Thr Phe Arg Val Asp Glu Asp  
145 150 155 160

Arg Phe Ile Ala Ser Ala Glu Gly Glu Phe Gly Asp Ala Gln Ile Glu  
165 170 175

Tyr Leu His Gly Glu Arg Ile Asp Lys Pro Ala Arg Ser Ile Tyr Ser  
180 185 190

Leu Asp Lys Ile Lys Glu Met Leu Lys Ala Asp Lys Phe Ser Glu Thr  
195 200 205

Ala Ile Ile Asn Leu Gly Asp Asp Met Pro Leu Lys Leu Thr Leu Lys  
210 215 220

Met Ala Ser Lys Glu Gly Glu Leu Ser Phe Leu Leu Ala Pro Arg Ile  
225 230 235 240

Glu Ala Glu Glu

(2) ANGABEN ZU SEQ ID NO: 16:

(i) SEQUENZKENNZEICHEN:

(A) LÄNGE: 469 Aminosäuren

(B) ART: Aminosäure

(C) STRANGFORM: Einzelstrang

(D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: Protein

(vi) URSPRÜNGLICHE HERKUNFT:

(A) ORGANISMUS: Homo sapiens

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 16:

Met Phe Ser Glu Gln Ala Ala Gln Arg Ala His Thr Leu Leu Ser Pro  
1 5 10 15

Pro Ser Ala Asn Asn Ala Thr Phe Ala Arg Val Pro Val Ala Thr Tyr  
20 25 30

Thr Asn Ser Ser Gln Pro Phe Arg Leu Gly Glu Arg Ser Phe Ser Arg  
35 40 45

Gln Tyr Ala His Ile Tyr Ala Thr Arg Leu Ile Gln Met Arg Pro Phe  
50 55 60

Leu Glu Asn Arg Ala Gln Gln His Trp Gly Ser Gly Val Gly Val Lys  
65 70 75 80

Lys Leu Cys Glu Leu Gln Pro Glu Glu Lys Cys Cys Val Val Gly Thr  
85 90 95

Leu Phe Lys Ala Met Pro Leu Gln Pro Ser Ile Leu Arg Glu Val Ser  
100 105 110

Glu Glu His Asn Leu Leu Pro Gln Pro Pro Arg Ser Lys Tyr Ile His  
115 120 125

Pro Asp Asp Glu Leu Val Leu Glu Asp Glu Leu Gln Arg Ile Lys Leu  
130 135 140

Lys Gly Thr Ile Asp Val Ser Lys Leu Val Thr Gly Thr Val Leu Ala  
145 150 155 160

Val Phe Gly Ser Val Arg Asp Asp Gly Lys Phe Leu Val Glu Asp Tyr  
165 170 175

Cys Phe Ala Asp Leu Ala Pro Gln Lys Pro Ala Pro Pro Leu Asp Thr  
180 185 190

# DE 198 40 771 A 1

Asp Arg Phe Val Leu Leu Val Ser Gly Leu Gly Leu Gly Gly Gly  
195 200 205

Gly Glu Ser Leu Leu Gly Thr Gln Leu Leu Val Asp Val Val Thr Gly  
210 215 220

Gln Leu Gly Asp Glu Gly Glu Gln Cys Ser Ala Ala His Val Ser Arg  
225 230 235 240

Val Ile Leu Ala Gly Asn Leu Leu Ser His Ser Thr Gln Ser Arg Asp  
245 250 255

Ser Ile Asn Lys Ala Lys Tyr Leu Thr Lys Lys Thr Gln Ala Ala Ser  
260 265 270

Val Glu Ala Val Lys Met Leu Asp Glu Ile Leu Leu Gln Leu Ser Ala  
275 280 285

Ser Val Pro Val Asp Val Met Pro Gly Glu Phe Asp Pro Thr Asn Tyr  
290 295 300

Thr Leu Pro Gln Gln Pro Leu His Pro Cys Met Phe Pro Leu Ala Thr  
305 310 315 320

Ala Tyr Ser Thr Leu Gln Leu Val Thr Asn Pro Tyr Gln Ala Thr Ile  
325 330 335

Asp Gly Val Arg Phe Leu Gly Thr Ser Gly Gln Asn Val Ser Asp Ile  
340 345 350

Phe Arg Tyr Ser Ser Met Glu Asp His Leu Glu Ile Leu Glu Trp Thr  
355 360 365

Leu Arg Val Arg His Ile Ser Pro Thr Ala Pro Asp Thr Leu Gly Cys  
370 375 380

Tyr Pro Phe Tyr Lys Thr Asp Pro Phe Ile Phe Pro Glu Cys Pro His  
385 390 395 400

Val Tyr Phe Cys Gly Asn Thr Pro Ser Phe Gly Ser Lys Ile Ile Arg  
405 410 415

Gly Pro Glu Asp Gln Thr Val Leu Leu Val Thr Val Pro Asp Phe Ser  
420 425 430

Ala Thr Gln Thr Ala Cys Leu Val Asn Leu Arg Ser Leu Ala Cys Gln  
435 440 445

Pro Ile Ser Phe Ser Gly Phe Gly Ala Glu Asp Asp Asp Leu Gly Gly  
450 455 460

Leu Gly Leu Gly Pro  
465

(2) ANGABEN ZU SEQ ID NO: 17:

(i) SEQUENZKENNZEICHEN:

(A) LÄNGE: 488 Aminosäuren

(B) ART: Aminosäure

(C) STRANGFORM: Einzelstrang

(D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: Protein

(vi) URSPRÜNGLICHE HERKUNFT:

(A) ORGANISMUS: Archaeoglobus fulgidus

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 17:

Met Val Ile Lys Asn Ile Asp Ala Ala Thr Val Ala Lys Lys Phe Leu  
1 5 10 15

Val Arg Gly Tyr Asn Ile Asp Pro Lys Ala Ala Glu Leu Ile Cys Lys  
20 25 30

Ser Gly Leu Phe Ser Asp Glu Leu Val Asp Lys Ile Cys Arg Ile Ala  
35 40 45

Asn Gly Gly Phe Ile Ile Glu Lys Ser Val Val Glu Glu Phe Leu Arg  
50 55 60

Asn Leu Ser Asn Leu Lys Pro Ala Thr Leu Thr Pro Arg Pro Glu Glu  
65 70 75 80

Arg Lys Val Glu Glu Val Lys Ala Ser Cys Ile Ala Leu Lys Val Ile  
85 90 95

Lys Asp Ile Thr Gly Lys Ser Ser Cys Gln Gly Asn Val Glu Asp Phe  
100 105 110

Leu Met Tyr Phe Asn Ser Arg Leu Glu Lys Leu Ser Arg Ile Ile Arg  
115 120 125



# DE 198 40 771 A 1

Ser Arg Val Asn Thr Thr Pro Ile Ala His Ala Gly Lys Val Arg Gly  
130 135 140

Asn Val Ser Val Val Gly Met Val Asn Glu Val Tyr Glu Arg Gly Asp  
145 150 155 160

Lys Cys Tyr Ile Arg Leu Glu Asp Thr Thr Gly Thr Ile Thr Cys Val  
165 170 175

Ala Thr Gly Lys Asn Ala Glu Val Ala Arg Glu Leu Leu Gly Asp Glu  
180 185 190

Val Ile Gly Val Thr Gly Leu Leu Lys Gly Ser Ser Leu Tyr Ala Asn  
195 200 205

Arg Ile Val Phe Pro Asp Val Pro Ile Asn Gly Asn Gly Glu Lys Lys  
210 215 220

Arg Asp Phe Tyr Ile Val Phe Leu Ser Asp Thr His Phe Gly Ser Lys  
225 230 235 240

Glu Phe Leu Glu Lys Glu Trp Glu Met Phe Val Arg Trp Leu Lys Gly  
245 250 255

Glu Val Gly Gly Lys Lys Ser Gln Asn Leu Ala Glu Lys Val Lys Tyr  
260 265 270

Ile Val Ile Ala Gly Asp Ile Val Asp Gly Ile Gly Val Tyr Pro Gly  
275 280 285

Gln Glu Asp Asp Leu Ala Ile Ser Asp Ile Tyr Gly Gln Tyr Glu Phe  
290 295 300

Ala Ala Ser His Leu Asp Glu Ile Pro Lys Glu Ile Lys Ile Ile Val  
305 310 315 320

Ser Pro Gly Asn His Asp Ala Val Arg Gln Ala Glu Pro Gln Pro Ala  
325 330 335

Phe Glu Gly Glu Ile Arg Ser Leu Phe Pro Lys Asn Val Glu His Val  
340 345 350

Gly Asn Pro Ala Tyr Val Asp Ile Glu Gly Val Lys Val Leu Ile Tyr  
355 360 365

His Gly Arg Ser Ile Asp Asp Ile Ile Ser Lys Ile Pro Arg Leu Ser  
370 375 380

# DE 198 40 771 A 1

Tyr Asp Glu Pro Gln Lys Val Met Glu Glu Leu Leu Lys Arg Arg His  
385 390 395 400

Leu Ser Pro Ile Tyr Gly Gly Arg Thr Pro Leu Ala Pro Glu Arg Glu  
405 410 415

Asp Tyr Leu Val Ile Glu Asp Val Pro Asp Ile Leu His Cys Gly His  
420 425 430

Ile His Thr Tyr Gly Thr Gly Phe Tyr Arg Gly Val Phe Met Val Asn  
435 440 445

Ser Ser Thr Trp Gln Ala Gln Thr Glu Phe Gln Lys Lys Val Asn Leu  
450 455 460

Asn Pro Met Pro Gly Asn Val Ala Val Tyr Arg Pro Gly Gly Glu Val  
465 470 475 480

Ile Arg Leu Arg Phe Tyr Gly Glu  
485

## (2) ANGABEN ZU SEQ ID NO: 18:

### (i) SEQUENZKENNZEICHEN:

(A) LÄNGE: 594 Aminosäuren

(B) ART: Aminosäure

(C) STRANGFORM: Einzelstrang

(D) TOPOLOGIE: linear

### (ii) ART DES MOLEKÜLS: Protein

### (vi) URSPRÜNGLICHE HERKUNFT:

(A) ORGANISMUS: Methanococcus jannaschii

### (xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 18:

Met Glu Ile Ile Asn Lys Phe Leu Asp Leu Glu Ala Leu Leu Ser Pro  
1 5 10 15

Thr Val Tyr Glu Lys Leu Lys Asn Phe Asp Glu Glu Lys Leu Lys Arg  
20 25 30

Leu Ile Gln Lys Ile Arg Glu Phe Lys Lys Tyr Asn Asn Ala Phe Ile  
35 40 45

# DE 198 40 771 A 1

Leu Leu Asp Glu Lys Phe Leu Asp Ile Phe Leu Gln Lys Asp Leu Asp  
 50 55 60

Glu Ile Ile Asn Glu Tyr Lys Asp Phe Asp Phe Ile Phe Tyr Tyr Thr  
 65 70 75 80

Gly Glu Glu Glu Lys Glu Lys Pro Lys Glu Val Lys Lys Glu Ile Lys  
 85 90 95

Lys Glu Thr Glu Glu Lys Ile Glu Lys Glu Lys Ile Glu Phe Val Lys  
 100 105 110

Lys Glu Glu Lys Glu Gln Phe Ile Lys Lys Ser Asp Glu Asp Val Glu  
 115 120 125

Glu Lys Leu Lys Gln Leu Ile Ser Lys Glu Glu Lys Lys Glu Asp Phe  
 130 135 140

Asp Ala Glu Arg Ala Lys Arg Tyr Glu His Ile Thr Lys Ile Lys Glu  
 145 150 155 160

Ser Val Asn Ser Arg Ile Lys Trp Ile Ala Lys Asp Ile Asp Ala Val  
 165 170 175

Ile Glu Ile Tyr Glu Asp Ser Asp Val Ser Gly Lys Ser Thr Cys Thr  
 180 185 190

Gly Thr Ile Glu Asp Phe Val Lys Tyr Phe Arg Asp Arg Phe Glu Arg  
 195 200 205

Leu Lys Val Phe Ile Glu Arg Lys Ala Gln Arg Lys Gly Tyr Pro Leu  
 210 215 220

Lys Asp Ile Lys Lys Met Lys Gly Gln Lys Asp Ile Phe Val Val Gly  
 225 230 235 240

Ile Val Ser Asp Val Asp Ser Thr Arg Asn Gly Asn Leu Ile Val Arg  
 245 250 255

Ile Glu Asp Thr Glu Asp Glu Ala Thr Leu Ile Leu Pro Lys Glu Lys  
 260 265 270

Ile Glu Ala Gly Lys Ile Pro Asp Asp Ile Leu Leu Asp Glu Val Ile  
 275 280 285

Gly Ala Ile Gly Thr Val Ser Lys Ser Gly Ser Ser Ile Tyr Val Asp  
 290 295 300

5  
 10  
 15  
 20  
 25  
 30  
 35  
 40  
 45  
 50  
 55  
 60  
 65

# DE 198 40 771 A 1

Glu Ile Ile Arg Pro Ala Leu Pro Pro Lys Glu Pro Lys Arg Ile Asp  
305 310 315 320

5 Glu Glu Ile Tyr Met Ala Phe Leu Ser Asp Ile His Val Gly Ser Lys  
325 330 335

10 Glu Phe Leu His Lys Glu Phe Glu Lys Phe Ile Arg Phe Leu Asn Gly  
340 345 350

15 Asp Val Asp Asn Glu Leu Glu Glu Lys Val Val Ser Arg Leu Lys Tyr  
355 360 365

20 Ile Cys Ile Ala Gly Asp Leu Val Asp Gly Val Gly Val Tyr Pro Gly  
370 375 380

Gln Glu Glu Asp Leu Tyr Glu Val Asp Ile Ile Glu Gln Tyr Arg Glu  
385 390 395 400

25 Ile Ala Met Tyr Leu Asp Gln Ile Pro Glu His Ile Ser Ile Ile Ile  
405 410 415

30 Ser Pro Gly Asn His Asp Ala Val Arg Pro Ala Glu Pro Gln Pro Lys  
420 425 430

35 Leu Pro Glu Lys Ile Thr Lys Leu Phe Asn Arg Asp Asn Ile Tyr Phe  
435 440 445

Val Gly Asn Pro Cys Thr Leu Asn Ile His Gly Phe Asp Thr Leu Leu  
450 455 460

40 Tyr His Gly Arg Ser Phe Asp Asp Leu Val Gly Gln Ile Arg Ala Ala  
465 470 475 480

45 Ser Tyr Glu Asn Pro Val Thr Ile Met Lys Glu Leu Ile Lys Arg Arg  
485 490 495

50 Leu Leu Cys Pro Thr Tyr Gly Gly Arg Cys Pro Ile Ala Pro Glu His  
500 505 510

Lys Asp Tyr Leu Val Ile Asp Arg Asp Ile Asp Ile Leu His Thr Gly  
515 520 525

55 His Ile His Ile Asn Gly Tyr Gly Ile Tyr Arg Gly Val Val Met Val  
530 535 540

60 Asn Ser Gly Thr Phe Gln Glu Gln Thr Asp Phe Gln Lys Arg Met Gly  
545 550 555 560

65

# DE 198 40 771 A 1

Ile Ser Pro Thr Pro Ala Ile Val Pro Ile Ile Asn Met Ala Lys Val  
565 570 575

Gly Glu Lys Gly His Tyr Leu Glu Trp Asp Arg Gly Val Leu Glu Val  
580 585 590

Arg Tyr

## (2) ANGABEN ZU SEQ ID NO: 19:

### (i) SEQUENZKENNZEICHEN:

- (A) LÄNGE: 622 Aminosäuren
- (B) ART: Aminosäure
- (C) STRANGFORM: Einzelstrang
- (D) TOPOLOGIE: linear

### (ii) ART DES MOLEKÜLS: Protein

### (vi) URSPRÜNGLICHE HERKUNFT:

- (A) ORGANISMUS: Pyrococcus horikoshii

### (xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 19:

Met Asp Glu Phe Val Lys Gly Leu Met Lys Asn Gly Tyr Leu Ile Thr  
1 5 10 15

Pro Ser Ala Tyr Tyr Leu Leu Val Gly His Phe Asn Glu Gly Lys Phe  
20 25 30

Ser Leu Ile Glu Leu Ile Lys Phe Ala Lys Ser Arg Glu Thr Phe Ile  
35 40 45

Ile Asp Asp Glu Ile Ala Asn Glu Phe Leu Lys Ser Ile Gly Ala Glu  
50 55 60

Val Glu Leu Pro Gln Glu Ile Lys Glu Gly Tyr Ile Ser Thr Gly Glu  
65 70 75 80

Gly Ser Gln Lys Val Pro Asp His Glu Glu Leu Glu Lys Ile Thr Asn  
85 90 95

Glu Ser Ser Val Glu Ser Ser Ile Ser Thr Gly Glu Thr Pro Lys Thr  
100 105 110

# DE 198 40 771 A 1

Glu Glu Leu Gln Pro Thr Leu Asp Ile Leu Glu Glu Ile Gly Asp  
115 120 125

5 Ile Glu Gly Gly Glu Ser Ser Ile Ser Thr Gly Asp Glu Val Pro Glu  
130 135 140

10 Val Glu Asn Asn Asn Gly Gly Thr Val Val Val Phe Asp Lys Tyr Gly  
145 150 155 160

15 Tyr Pro Phe Thr Tyr Val Pro Glu Glu Ile Glu Glu Glu Leu Glu Glu  
165 170 175

Tyr Pro Lys Tyr Glu Asp Val Thr Ile Glu Ile Asn Pro Asn Leu Glu  
180 185 190

20 Val Val Pro Ile Glu Lys Asp Tyr Glu Ile Lys Phe Asp Val Arg Arg  
195 200 205

25 Val Lys Leu Lys Pro Pro Lys Val Lys Ser Gly Ser Gly Lys Glu Gly  
210 215 220

30 Glu Ile Ile Val Glu Ala Tyr Ala Ser Leu Phe Arg Ser Arg Leu Arg  
225 230 235 240

35 Lys Leu Arg Arg Ile Leu Arg Glu Asn Pro Glu Val Ser Asn Val Ile  
245 250 255

40 Asp Ile Lys Lys Leu Lys Tyr Val Lys Gly Asp Glu Glu Val Thr Ile  
260 265 270

Ile Gly Leu Val Asn Ser Lys Lys Glu Thr Ser Lys Gly Leu Ile Phe  
275 280 285

45 Glu Val Glu Asp Gln Thr Asp Arg Val Lys Val Phe Leu Pro Lys Asp  
290 295 300

50 Ser Glu Asp Tyr Arg Glu Ala Leu Lys Val Leu Pro Asp Ala Val Val  
305 310 315 320

55 Ala Phe Lys Gly Val Tyr Ser Lys Arg Gly Ile Phe Phe Ala Asn Arg  
325 330 335

Phe Tyr Leu Pro Asp Val Pro Leu Tyr Arg Lys Gln Lys Pro Pro Leu  
340 345 350

60 Glu Glu Lys Val Tyr Ala Val Leu Thr Ser Asp Ile His Val Gly Ser  
355 360 365

65

# DE 198 40 771 A 1

Lys Glu Phe Cys Glu Lys Ala Phe Ile Lys Phe Leu Glu Trp Leu Asn  
 370 375 380  
 Gly Tyr Val Glu Ser Lys Glu Glu Glu Ile Val Ser Arg Ile Arg  
 385 390 395 400  
 Tyr Leu Ile Ile Ala Gly Asp Val Val Asp Gly Ile Gly Ile Tyr Pro  
 405 410 415  
 Gly Gln Tyr Ser Asp Leu Ile Ile Pro Asp Ile Phe Asp Gln Tyr Glu  
 420 425 430  
 Ala Leu Ala Asn Leu Leu Ser Asn Val Pro Lys His Ile Thr Ile Phe  
 435 440 445  
 Ile Gly Pro Gly Asn His Asp Ala Ala Arg Pro Ala Ile Pro Gln Pro  
 450 455 460  
 Glu Phe Tyr Glu Glu Tyr Ala Lys Pro Leu Tyr Lys Leu Lys Asn Thr  
 465 470 475 480  
 Val Ile Ile Ser Asn Pro Ala Val Ile Arg Leu His Gly Arg Asp Phe  
 485 490 495  
 Leu Ile Ala His Gly Arg Gly Ile Glu Asp Val Val Ser Phe Val Pro  
 500 505 510  
 Gly Leu Thr His His Lys Pro Gly Leu Pro Met Val Glu Leu Leu Lys  
 515 520 525  
 Met Arg His Leu Ala Pro Thr Phe Gly Gly Lys Val Pro Ile Ala Pro  
 530 535 540  
 Asp Pro Glu Asp Leu Leu Val Ile Glu Glu Val Pro Asp Leu Val Gln  
 545 550 555 560  
 Met Gly His Val His Val Tyr Asp Thr Ala Val Tyr Arg Gly Val Gln  
 565 570 575  
 Leu Val Asn Ser Ala Thr Trp Gln Ala Gln Thr Glu Phe Gln Lys Met  
 580 585 590  
 Val Asn Ile Val Pro Thr Pro Gly Leu Val Pro Ile Val Asp Val Glu  
 595 600 605  
 Ser Ala Arg Val Ile Lys Val Leu Asp Phe Ser Arg Trp Cys  
 610 615 620

5  
 10  
 15  
 20  
 25  
 30  
 35  
 40  
 45  
 50  
 55  
 60  
 65

## (2) ANGABEN ZU SEQ ID NO: 20:

## (i) SEQUENZKENNZEICHEN:

- (A) LÄNGE: 482 Aminosäuren  
 (B) ART: Aminosäure  
 (C) STRANGFORM: Einzelstrang  
 (D) TOPOLOGIE: linear

## (ii) ART DES MOLEKÜLS: Protein

## (vi) URSPRÜNGLICHE HERKUNFT:

- (A) ORGANISMUS: *Methanobacterium thermoautotrophicum*

## (xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 20:

Met Asn Glu Ile Ile Gly Lys Phe Ala Arg Glu Gly Ile Leu Ile Glu  
 1 5 10 15

Asp Asn Ala Tyr Phe Arg Leu Arg Glu Met Asp Asp Pro Ala Ser Val  
 20 25 30

Ser Ser Glu Leu Ile Val Lys Ile Lys Ser Asn Gly Gly Lys Phe Thr  
 35 40 45

Val Leu Thr Ser Glu Met Leu Asp Glu Phe Phe Glu Ile Asp Asn Pro  
 50 55 60

Ala Glu Ile Lys Ala Arg Gly Pro Leu Met Val Pro Ala Glu Arg Asp  
 65 70 75 80

Phe Asp Phe Glu Val Ile Ser Asp Thr Ser Asn Arg Ser Tyr Thr Ser  
 85 90 95

Gly Glu Ile Gly Asp Met Ile Ala Tyr Phe Asn Ser Arg Tyr Ser Ser  
 100 105 110

Leu Lys Asn Leu Leu Ser Lys Arg Pro Glu Leu Lys Gly His Ile Pro  
 115 120 125

Ile Ala Asp Leu Arg Gly Gly Glu Asp Val Val Ser Ile Ile Gly Met  
 130 135 140

Val Asn Asp Val Arg Asn Thr Lys Asn Asn His Arg Ile Ile Glu Leu  
 145 150 155 160



# DE 198 40 771 A 1

Glu Asp Asp Thr Gly Glu Ile Ser Val Val Val His Asn Glu Asn His  
 165 170 175  
 Lys Leu Phe Glu Lys Ser Glu Lys Ile Val Arg Asp Glu Val Val Gly  
 180 185 190  
 Val His Gly Thr Lys Lys Gly Arg Phe Val Val Ala Ser Glu Ile Phe  
 195 200 205  
 His Pro Gly Val Pro Arg Ile Gln Glu Lys Glu Met Asp Phe Ser Val  
 210 215 220  
 Ala Phe Ile Ser Asp Val His Ile Gly Ser Gln Thr Phe Leu Glu Asp  
 225 230 235 240  
 Ala Phe Met Lys Phe Val Lys Trp Ile Asn Gly Asp Phe Gly Ser Glu  
 245 250 255  
 Glu Gln Arg Ser Leu Ala Ala Asp Val Lys Tyr Leu Val Val Ala Gly  
 260 265 270  
 Asp Ile Val Asp Gly Ile Gly Ile Tyr Pro Gly Gln Glu Lys Glu Leu  
 275 280 285  
 Leu Ile Arg Asp Ile His Glu Gln Tyr Glu Glu Ala Ala Arg Leu Phe  
 290 295 300  
 Gly Asp Ile Arg Ser Asp Ile Lys Ile Val Met Ile Pro Gly Asn His  
 305 310 315 320  
 Asp Ser Ser Arg Ile Ala Glu Pro Gln Pro Ala Ile Pro Glu Glu Tyr  
 325 330 335  
 Ala Lys Ser Leu Tyr Ser Ile Arg Asn Ile Glu Phe Leu Ser Asn Pro  
 340 345 350  
 Ser Leu Val Ser Leu Asp Gly Val Arg Thr Leu Ile Tyr His Gly Arg  
 355 360 365  
 Ser Phe Asp Asp Met Ala Met Ser Val Asn Gly Leu Ser His Glu Arg  
 370 375 380  
 Ser Asp Leu Ile Met Glu Glu Leu Leu Glu Lys Arg His Leu Ala Pro  
 385 390 395 400  
 Ile Tyr Gly Glu Arg Thr Pro Leu Ala Ser Glu Ile Glu Asp His Leu  
 405 410 415

5

10

15

20

25

30

35

40

45

50

55

60

65

# DE 198 40 771 A 1

Val Ile Asp Glu Val Pro His Val Leu His Thr Gly His Val His Ile  
420 425 430

Asn Ala Tyr Lys Lys Tyr Lys Gly Val His Leu Ile Asn Ser Gly Thr  
435 440 445

Phe Gln Ser Gln Thr Glu Phe Gln Lys Ile Tyr Asn Ile Val Pro Thr  
450 455 460

Cys Gly Gln Val Pro Val Leu Asn Arg Gly Val Met Lys Leu Leu Glu  
465 470 475 480

Phe Ser

## (2) ANGABEN ZU SEQ ID NO: 21:

### (i) SEQUENZKENNZEICHEN:

- (A) LÄNGE: 613 Aminosäuren
- (B) ART: Aminosäure
- (C) STRANGFORM: Einzelstrang
- (D) TOPOLOGIE: linear

### (ii) ART DES MOLEKÜLS: Protein

### (vi) URSPRÜNGLICHE HERKUNFT:

- (A) ORGANISMUS: Pyrococcus furiosus

### (xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 21:

Met Asp Glu Phe Val Lys Ser Leu Leu Lys Ala Asn Tyr Leu Ile Thr  
1 5 10 15

Pro Ser Ala Tyr Tyr Leu Leu Arg Glu Tyr Tyr Glu Lys Gly Glu Phe  
20 25 30

Ser Ile Val Glu Leu Val Lys Phe Ala Arg Ser Arg Glu Ser Tyr Ile  
35 40 45

Ile Thr Asp Ala Leu Ala Thr Glu Phe Leu Lys Val Lys Gly Leu Glu  
50 55 60

Pro Ile Leu Pro Val Glu Thr Lys Gly Gly Phe Val Ser Thr Gly Glu  
65 70 75 80

# DE 198 40 771 A 1

Ser Gln Lys Glu Gln Ser Tyr Glu Glu Ser Phe Gly Thr Lys Glu Glu	5
85 90 95	
Ile Ser Gln Glu Ile Lys Glu Gly Glu Ser Phe Ile Ser Thr Gly Ser	10
100 105 110	
Glu Pro Leu Glu Glu Glu Leu Asn Ser Ile Gly Ile Glu Glu Ile Gly	15
115 120 125	
Ala Asn Glu Glu Leu Val Ser Asn Gly Asn Asp Asn Gly Gly Glu Ala	20
130 135 140	
Ile Val Phe Asp Lys Tyr Gly Tyr Pro Met Val Tyr Ala Pro Glu Glu	25
145 150 155 160	
Ile Glu Val Glu Glu Lys Glu Tyr Ser Lys Tyr Glu Asp Leu Thr Ile	30
165 170 175	
Pro Met Asn Pro Asp Phe Asn Tyr Val Glu Ile Lys Glu Asp Tyr Asp	35
180 185 190	
Val Val Phe Asp Val Arg Asn Val Lys Leu Lys Pro Pro Lys Val Lys	40
195 200 205	
Asn Gly Asn Gly Lys Glu Gly Glu Ile Ile Val Glu Ala Tyr Ala Ser	45
210 215 220	
Leu Phe Arg Ser Arg Leu Lys Lys Leu Arg Lys Ile Leu Arg Glu Asn	50
225 230 235 240	
Pro Glu Leu Asp Asn Val Val Asp Ile Gly Lys Leu Lys Tyr Val Lys	55
245 250 255	
Glu Asp Glu Thr Val Thr Ile Ile Gly Leu Val Asn Ser Lys Arg Glu	60
260 265 270	
Val Asn Lys Gly Leu Ile Phe Glu Ile Glu Asp Leu Thr Gly Lys Val	65
275 280 285	
Lys Val Phe Leu Pro Lys Asp Ser Glu Asp Tyr Arg Glu Ala Phe Lys	
290 295 300	
Val Leu Pro Asp Ala Val Val Ala Phe Lys Gly Val Tyr Ser Lys Arg	
305 310 315 320	
Gly Ile Leu Tyr Ala Asn Lys Phe Tyr Leu Pro Asp Val Pro Leu Tyr	
325 330 335	

# DE 198 40 771 A 1

Arg Arg Gln Lys Pro Pro Leu Glu Glu Lys Val Tyr Ala Ile Leu Ile  
 340 345 350  
 5 Ser Asp Ile His Val Gly Ser Lys Glu Phe Cys Glu Asn Ala Phe Ile  
 355 360 365  
 10 Lys Phe Leu Glu Trp Leu Asn Gly Asn Val Glu Thr Lys Glu Glu Glu  
 370 375 380  
 Glu Ile Val Ser Arg Val Lys Tyr Leu Ile Ile Ala Gly Asp Val Val  
 15 385 390 395 400  
 Asp Gly Val Gly Val Tyr Pro Gly Gln Tyr Ala Asp Leu Thr Ile Pro  
 20 405 410 415  
 Asp Ile Phe Asp Gln Tyr Glu Ala Leu Ala Asn Leu Leu Ser His Val  
 25 420 425 430  
 Pro Lys His Ile Thr Met Phe Ile Ala Pro Gly Asn His Asp Ala Ala  
 30 435 440 445  
 Arg Gln Ala Ile Pro Gln Pro Glu Phe Tyr Lys Glu Tyr Ala Lys Pro  
 35 450 455 460  
 Ile Tyr Lys Leu Lys Asn Ala Val Ile Ile Ser Asn Pro Ala Val Ile  
 40 465 470 475 480  
 Arg Leu His Gly Arg Asp Phe Leu Ile Ala His Gly Arg Gly Ile Glu  
 45 485 490 495  
 Asp Val Val Gly Ser Val Pro Gly Leu Thr His His Lys Pro Gly Leu  
 50 500 505 510  
 Pro Met Val Glu Leu Leu Lys Met Arg His Val Ala Pro Met Phe Gly  
 55 515 520 525  
 Gly Lys Val Pro Ile Ala Pro Asp Pro Glu Asp Leu Leu Val Ile Glu  
 60 530 535 540  
 Glu Val Pro Asp Val Val His Met Gly His Val His Val Tyr Asp Ala  
 65 545 550 555 560  
 Val Val Tyr Arg Gly Val Gln Leu Val Asn Ser Ala Thr Trp Gln Ala  
 565 570 575  
 Gln Thr Glu Phe Gln Lys Met Val Asn Ile Val Pro Thr Pro Ala Lys  
 580 585 590

Val Pro Val Val Asp Ile Asp Thr Ala Lys Val Val Lys Val Leu Asp  
 595 600 605

Phe Ser Gly Trp Cys  
 610

(2) ANGABEN ZU SEQ ID NO: 22:

(i) SEQUENZKENNZEICHEN:

- (A) LÄNGE: 1107 Aminosäuren
- (B) ART: Aminosäure
- (C) STRANGFORM: Einzelstrang
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: Protein

(vi) URSPRÜNGLICHE HERKUNFT:

- (A) ORGANISMUS: Homo sapiens

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 22:

Met Asp Gly Lys Arg Arg Pro Gly Pro Gly Pro Gly Val Pro Pro Lys  
 1 5 10 15

Arg Ala Arg Gly Gly Leu Trp Asp Asp Asp Ala Pro Trp Pro Ser  
 20 25 30

Gln Phe Glu Glu Asp Leu Ala Leu Met Glu Glu Met Glu Ala Glu His  
 35 40 45

Arg Leu Gln Glu Gln Glu Glu Glu Leu Gln Ser Val Leu Glu Gly  
 50 55 60

Val Ala Asp Gly Gln Val Pro Pro Ser Ala Ile Asp Pro Arg Trp Leu  
 65 70 75 80

Arg Pro Thr Pro Pro Ala Leu Asp Pro Gln Thr Glu Pro Leu Ile Phe  
 85 90 95

Gln Gln Leu Glu Ile Asp His Tyr Val Gly Pro Ala Gln Pro Val Pro  
 100 105 110

Gly Gly Pro Pro Pro Ser Arg Gly Ser Val Pro Val Leu Arg Ala Phe  
 115 120 125

# DE 198 40 771 A 1

Gly Val Thr Asp Glu Gly Phe Ser Val Cys Cys His Ile His Gly Phe  
 130 135 140

5 Ala Pro Tyr Phe Tyr Thr Pro Ala Pro Pro Gly Phe Gly Pro Glu His  
 145 150 155 160

10 Met Gly Asp Leu Gln Arg Glu Leu Asn Leu Ala Ile Ser Arg Asp Ser  
 165 170 175

15 Arg Gly Gly Arg Glu Leu Thr Gly Pro Ala Val Leu Ala Val Glu Leu  
 180 185 190

20 Cys Ser Arg Glu Ser Met Phe Gly Tyr His Gly His Gly Pro Ser Pro  
 195 200 205

25 Phe Leu Arg Ile Thr Val Ala Leu Pro Arg Leu Val Ala Pro Ala Arg  
 210 215 220

30 Arg Leu Leu Glu Gln Gly Ile Arg Val Ala Gly Leu Gly Thr Pro Ser  
 225 230 235 240

35 Phe Ala Pro Tyr Glu Ala Asn Val Asp Phe Glu Ile Arg Phe Met Val  
 245 250 255

40 Asp Thr Asp Ile Val Gly Cys Asn Trp Leu Glu Leu Pro Ala Gly Lys  
 260 265 270

45 Tyr Ala Leu Arg Leu Lys Glu Lys Ala Thr Gln Cys Gln Leu Glu Ala  
 275 280 285

50 Asp Val Leu Trp Ser Asp Val Val Ser His Pro Pro Glu Gly Pro Trp  
 290 295 300

55 Gln Arg Ile Ala Pro Leu Arg Val Leu Ser Phe Asp Ile Glu Cys Ala  
 305 310 315 320

60 Gly Arg Lys Gly Ile Phe Pro Glu Pro Glu Arg Asp Pro Val Ile Gln  
 325 330 335

65 Ile Cys Ser Leu Gly Leu Arg Trp Gly Glu Pro Glu Pro Phe Leu Arg  
 340 345 350

70 Leu Ala Leu Thr Leu Arg Pro Cys Ala Pro Ile Leu Gly Ala Lys Val  
 355 360 365

75 Gln Ser Tyr Glu Lys Glu Glu Asp Leu Leu Gln Ala Trp Ser Thr Phe  
 370 375 380

# DE 198 40 771 A 1

Ile Arg Ile Met Asp Pro Asp Val Ile Thr Gly Tyr Asn Ile Gln Asn	
385                      390                      395                      400	
Phe Asp Leu Pro Tyr Leu Ile Ser Arg Ala Gln Thr Leu Lys Val Gln	5
405                      410                      415	
Thr Phe Pro Phe Leu Gly Arg Val Ala Gly Leu Cys Ser Asn Ile Arg	10
420                      425                      430	
Asp Ser Ser Phe Gln Ser Lys Gln Thr Gly Arg Arg Asp Thr Lys Val	15
435                      440                      445	
Val Ser Met Val Gly Arg Val Gln Met Asp Met Leu Gln Val Leu Leu	20
450                      455                      460	
Arg Glu Tyr Lys Leu Arg Ser His Thr Leu Asn Ala Val Ser Phe His	25
465                      470                      475                      480	
Phe Leu Gly Glu Gln Lys Glu Asp Val Gln His Ser Ile Ile Thr Asp	30
485                      490                      495	
Leu Gln Asn Gly Asn Asp Gln Thr Arg Arg Arg Leu Ala Val Tyr Cys	35
500                      505                      510	
Leu Lys Asp Ala Tyr Leu Pro Leu Arg Leu Leu Glu Arg Leu Met Val	40
515                      520                      525	
Leu Val Asn Ala Val Glu Met Ala Arg Val Thr Gly Val Pro Leu Ser	45
530                      535                      540	
Tyr Leu Leu Ser Arg Gly Gln Gln Val Lys Val Val Ser Gln Leu Leu	50
545                      550                      555                      560	
Arg Gln Ala Met His Glu Gly Leu Leu Met Pro Val Val Lys Ser Glu	55
565                      570                      575	
Gly Gly Glu Asp Tyr Thr Gly Ala Thr Val Ile Glu Pro Leu Lys Gly	60
580                      585                      590	
Tyr Tyr Asp Val Pro Ile Ala Thr Leu Asp Phe Ser Ser Leu Tyr Pro	65
595                      600                      605	
Ser Ile Met Met Ala His Asn Leu Cys Tyr Thr Thr Leu Leu Arg Pro	
610                      615                      620	
Gly Thr Ala Gln Lys Leu Gly Leu Thr Glu Asp Gln Phe Ile Arg Thr	
625                      630                      635                      640	

# DE 198 40 771 A 1

Pro Thr Gly Asp Glu Phe Val Lys Thr Ser Val Arg Lys Gly Leu Leu  
 645 650 655

5 Pro Gln Ile Leu Glu Asn Leu Leu Ser Ala Arg Lys Arg Ala Lys Ala  
 660 665 670

10 Glu Leu Ala Lys Glu Thr Asp Pro Leu Arg Arg Gln Val Leu Asp Gly  
 675 680 685

15 Arg Gln Leu Ala Leu Lys Val Ser Ala Asn Ser Val Tyr Gly Phe Thr  
 690 695 700

20 Gly Ala Gln Val Gly Lys Leu Pro Cys Leu Glu Ile Ser Gln Ser Val  
 705 710 715 720

Thr Gly Phe Gly Arg Gln Met Ile Glu Lys Thr Lys Gln Leu Val Glu  
 725 730 735

25 Ser Lys Tyr Thr Val Glu Asn Gly Tyr Ser Thr Ser Ala Lys Val Val  
 740 745 750

30 Tyr Gly Asp Thr Asp Ser Val Met Cys Arg Phe Gly Val Ser Ser Val  
 755 760 765

35 Ala Glu Ala Met Ala Leu Gly Arg Glu Ala Ala Asp Trp Val Ser Gly  
 770 775 780

His Phe Pro Ser Pro Ile Arg Leu Glu Phe Glu Lys Val Tyr Phe Pro  
 785 790 795 800

40 Tyr Leu Leu Ile Ser Lys Lys Arg Tyr Ala Gly Leu Leu Phe Ser Ser  
 805 810 815

45 Arg Pro Asp Ala His Asp Arg Met Asp Cys Lys Gly Leu Glu Ala Val  
 820 825 830

50 Arg Arg Asp Asn Cys Pro Leu Val Ala Asn Leu Val Thr Ala Ser Leu  
 835 840 845

55 Arg Arg Leu Leu Ile Asp Arg Asp Pro Glu Gly Ala Val Ala His Ala  
 850 855 860

Gln Asp Val Ile Ser Asp Leu Leu Cys Asn Arg Ile Asp Ile Ser Gln  
 865 870 875 880

60 Leu Val Ile Thr Lys Glu Leu Thr Arg Ala Ala Ser Asp Tyr Ala Gly  
 885 890 895

65



# DE 198 40 771 A 1

Lys Gln Ala His Val Glu Leu Ala Glu Arg Met Arg Lys Arg Asp Pro  
900 905 910

Gly Ser Ala Pro Ser Leu Gly Asp Arg Val Pro Tyr Val Ile Ile Ser  
915 920 925

Ala Ala Lys Gly Val Ala Ala Tyr Met Lys Ser Glu Asp Pro Leu Phe  
930 935 940

Val Leu Glu His Ser Leu Pro Ile Asp Thr Gln Tyr Tyr Leu Glu Gln  
945 950 955 960

Gln Leu Ala Lys Pro Leu Leu Arg Ile Phe Glu Pro Ile Leu Gly Glu  
965 970 975

Gly Arg Ala Glu Ala Val Leu Leu Arg Gly Asp His Thr Arg Cys Lys  
980 985 990

Thr Val Leu Thr Gly Lys Val Gly Gly Leu Leu Ala Phe Ala Lys Arg  
995 1000 1005

Arg Asn Cys Cys Ile Gly Cys Arg Thr Val Leu Ser His Gln Gly Ala  
1010 1015 1020

Val Cys Glu Phe Cys Gln Pro Arg Glu Ser Glu Leu Tyr Gln Lys Glu  
1025 1030 1035 1040

Val Ser His Leu Asn Ala Leu Glu Glu Arg Phe Ser Arg Leu Trp Thr  
1045 1050 1055

Gln Cys Gln Arg Cys Gln Gly Ser Leu His Glu Asp Val Ile Cys Thr  
1060 1065 1070

Ser Arg Asp Cys Pro Ile Phe Tyr Met Arg Lys Lys Val Arg Lys Asp  
1075 1080 1085

Leu Glu Asp Gln Glu Gln Leu Leu Arg Arg Phe Gly Pro Pro Gly Pro  
1090 1095 1100

Glu Ala Trp  
1105

(2) ANGABEN ZU SEQ ID NO: 23:

(i) SEQUENZKENNZEICHEN:

(A) LÄNGE: 781 Aminosäuren

(B) ART: Aminosäure

(C) STRANGFORM: Einzelstrang

(D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: Protein

(vi) URSPRÜNGLICHE HERKUNFT:

(A) ORGANISMUS: Archaeoglobus fulgidus

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 23:

Met Glu Arg Val Glu Gly Trp Leu Ile Asp Ala Asp Tyr Glu Thr Ile  
 1 5 10 15

Gly Gly Lys Ala Val Val Arg Leu Trp Cys Lys Asp Asp Gln Gly Ile  
 20 25 30

Phe Val Ala Tyr Asp Tyr Asn Phe Asp Pro Tyr Phe Tyr Val Ile Gly  
 35 40 45

Val Asp Glu Asp Ile Leu Lys Asn Ala Ala Thr Ser Thr Arg Arg Glu  
 50 55 60

Val Ile Lys Leu Lys Ser Phe Glu Lys Ala Gln Leu Lys Thr Leu Gly  
 65 70 75 80

Arg Glu Val Glu Gly Tyr Ile Val Tyr Ala His His Pro Gln His Val  
 85 90 95

Pro Lys Leu Arg Asp Tyr Leu Ser Gln Phe Gly Asp Val Arg Glu Ala  
 100 105 110

Asp Ile Pro Phe Ala Tyr Arg Tyr Leu Ile Asp Lys Asp Leu Ala Cys  
 115 120 125

Met Asp Gly Ile Ala Ile Glu Gly Glu Lys Gln Gly Gly Val Ile Arg  
 130 135 140

Ser Tyr Lys Ile Glu Lys Val Glu Arg Ile Pro Arg Met Glu Phe Pro  
 145 150 155 160

Glu Leu Lys Met Leu Val Phe Asp Cys Glu Met Leu Ser Ser Phe Gly  
 165 170 175

Met Pro Glu Pro Glu Lys Asp Pro Ile Ile Val Ile Ser Val Lys Thr  
 180 185 190

# DE 198 40 771 A 1

Asn Asp Asp Asp Glu Ile Ile Leu Thr Gly Asp Glu Arg Lys Ile Ile	
195 200 205	
Ser Asp Phe Val Lys Leu Ile Lys Ser Tyr Asp Pro Asp Ile Ile Val	5
210 215 220	
Gly Tyr Asn Gln Asp Ala Phe Asp Trp Pro Tyr Leu Arg Lys Arg Ala	10
225 230 235 240	
Glu Arg Trp Asn Ile Pro Leu Asp Val Gly Arg Asp Gly Ser Asn Val	15
245 250 255	
Val Phe Arg Gly Gly Arg Pro Lys Ile Thr Gly Arg Leu Asn Val Asp	20
260 265 270	
Leu Tyr Asp Ile Ala Met Arg Ile Ser Asp Ile Lys Ile Lys Lys Leu	25
275 280 285	
Glu Asn Val Ala Glu Phe Leu Gly Thr Lys Ile Glu Ile Ala Asp Ile	30
290 295 300	
Glu Ala Lys Asp Ile Tyr Arg Tyr Trp Ser Arg Gly Glu Lys Glu Lys	35
305 310 315 320	
Val Leu Asn Tyr Ala Arg Gln Asp Ala Ile Asn Thr Tyr Leu Ile Ala	40
325 330 335	
Lys Glu Leu Leu Pro Met His Tyr Glu Leu Ser Lys Met Ile Arg Leu	45
340 345 350	
Pro Val Asp Asp Val Thr Arg Met Gly Arg Gly Lys Gln Val Asp Trp	50
355 360 365	
Leu Leu Leu Ser Glu Ala Lys Lys Ile Gly Glu Ile Ala Pro Asn Pro	55
370 375 380	
Pro Glu His Ala Glu Ser Tyr Glu Gly Ala Phe Val Leu Glu Pro Glu	60
385 390 395 400	
Arg Gly Leu His Glu Asn Val Ala Cys Leu Asp Phe Ala Ser Met Tyr	65
405 410 415	
Pro Ser Ile Met Ile Ala Phe Asn Ile Ser Pro Asp Thr Tyr Gly Cys	
420 425 430	
Arg Asp Asp Cys Tyr Glu Ala Pro Glu Val Gly His Lys Phe Arg Lys	
435 440 445	

# DE 198 40 771 A 1

Ser Pro Asp Gly Phe Phe Lys Arg Ile Leu Arg Met Leu Ile Glu Lys  
450 455 460

5 Arg Arg Glu Leu Lys Val Glu Leu Lys Asn Leu Ser Pro Glu Ser Ser  
465 470 475 480

10 Glu Tyr Lys Leu Leu Asp Ile Lys Gln Gln Thr Leu Lys Val Leu Thr  
485 490 495

15 Asn Ser Phe Tyr Gly Tyr Met Gly Trp Asn Leu Ala Arg Trp Tyr Cys  
500 505 510

20 His Pro Cys Ala Glu Ala Thr Thr Ala Trp Gly Arg His Phe Ile Arg  
515 520 525

Thr Ser Ala Lys Ile Ala Glu Ser Met Gly Phe Lys Val Leu Tyr Gly  
530 535 540

25 Asp Thr Asp Ser Ile Phe Val Thr Lys Ala Gly Met Thr Lys Glu Asp  
545 550 555 560

30 Val Asp Arg Leu Ile Asp Lys Leu His Glu Glu Leu Pro Ile Gln Ile  
565 570 575

35 Glu Val Asp Glu Tyr Tyr Ser Ala Ile Phe Phe Val Glu Lys Lys Arg  
580 585 590

Tyr Ala Gly Leu Thr Glu Asp Gly Arg Leu Val Val Lys Gly Leu Glu  
595 600 605

40 Val Arg Arg Gly Asp Trp Cys Glu Leu Ala Lys Lys Val Gln Arg Glu  
610 615 620

45 Val Ile Glu Val Ile Leu Lys Glu Lys Asn Pro Glu Lys Ala Leu Ser  
625 630 635 640

50 Leu Val Lys Asp Val Ile Leu Arg Ile Lys Glu Gly Lys Val Ser Leu  
645 650 655

55 Glu Glu Val Val Ile Tyr Lys Gly Leu Thr Lys Lys Pro Ser Lys Tyr  
660 665 670

Glu Ser Met Gln Ala His Val Lys Ala Ala Leu Lys Ala Arg Glu Met  
675 680 685

60 Gly Ile Ile Tyr Pro Val Ser Ser Lys Ile Gly Tyr Val Ile Val Lys  
690 695 700

65

Gly Ser Gly Asn Ile Gly Asp Arg Ala Tyr Pro Ile Asp Leu Ile Glu  
705 710 715 720

Asp Phe Asp Gly Glu Asn Leu Arg Ile Lys Thr Lys Ser Gly Ile Glu  
725 730 735

Ile Lys Lys Leu Asp Lys Asp Tyr Tyr Ile Asp Asn Gln Ile Ile Pro  
740 745 750

Ser Val Leu Arg Ile Leu Glu Arg Phe Gly Tyr Thr Glu Ala Ser Leu  
755 760 765

Lys Gly Ser Ser Gln Met Ser Leu Asp Ser Phe Phe Ser  
770 775 780

(2) ANGABEN ZU SEQ ID NO: 24:

(i) SEQUENZKENNZEICHEN:

- (A) LÄNGE: 1634 Aminosäuren
- (B) ART: Aminosäure
- (C) STRANGFORM: Einzelstrang
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: Protein

(vi) URSPRÜNGLICHE HERKUNFT:

- (A) ORGANISMUS: Methanococcus jannaschii

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 24:

Met Gly Met Ser Met Gly Lys Ile Lys Ile Asp Ala Leu Ile Asp Asn  
1 5 10 15

Thr Tyr Lys Thr Ile Glu Asp Lys Ala Val Ile Tyr Leu Tyr Leu Ile  
20 25 30

Asn Ser Ile Leu Lys Asp Arg Asp Phe Lys Pro Tyr Phe Tyr Val Glu  
35 40 45

Leu His Lys Glu Lys Val Glu Asn Glu Asp Ile Glu Lys Ile Lys Glu  
50 55 60

Phe Leu Leu Lys Asn Asp Leu Leu Lys Phe Val Glu Asn Ile Glu Val  
65 70 75 80

# DE 198 40 771 A 1

Val Lys Lys Ile Ile Leu Arg Lys Glu Lys Glu Val Ile Lys Ile Ile  
85 90 95

Ala Thr His Pro Gln Lys Val Pro Lys Leu Arg Lys Ile Lys Glu Cys  
100 105 110

Glu Ile Val Lys Glu Ile Tyr Glu His Asp Ile Pro Phe Ala Lys Arg  
115 120 125

Tyr Leu Ile Asp Asn Glu Ile Ile Pro Met Thr Tyr Trp Asp Phe Glu  
130 135 140

Asn Lys Lys Pro Val Ser Ile Glu Ile Pro Lys Leu Lys Ser Val Ala  
145 150 155 160

Phe Asp Met Glu Val Tyr Asn Arg Asp Thr Glu Pro Asn Pro Glu Arg  
165 170 175

Asp Pro Ile Leu Met Ala Ser Phe Trp Asp Glu Asn Gly Gly Lys Val  
180 185 190

Ile Thr Tyr Lys Glu Phe Asn His Pro Asn Ile Glu Val Val Lys Asn  
195 200 205

Glu Lys Glu Leu Ile Lys Lys Ile Ile Glu Thr Leu Lys Glu Tyr Asp  
210 215 220

Val Ile Tyr Thr Tyr Asn Gly Asp Asn Phe Asp Phe Pro Tyr Leu Lys  
225 230 235 240

Ala Arg Ala Lys Ile Tyr Gly Ile Asp Ile Asn Leu Gly Lys Asp Gly  
245 250 255

Glu Glu Leu Lys Ile Lys Arg Gly Gly Met Glu Tyr Arg Ser Tyr Ile  
260 265 270

Pro Gly Arg Val His Ile Asp Leu Tyr Pro Ile Ser Arg Arg Leu Leu  
275 280 285

Lys Leu Thr Lys Tyr Thr Leu Glu Asp Val Val Tyr Asn Leu Phe Gly  
290 295 300

Ile Glu Lys Leu Lys Ile Pro His Thr Lys Ile Val Asp Tyr Trp Ala  
305 310 315 320

Asn Asn Asp Lys Thr Leu Ile Glu Tyr Ser Leu Gln Asp Ala Lys Tyr  
325 330 335

# DE 198 40 771 A 1

Thr Tyr Lys Ile Gly Lys Tyr Phe Phe Pro Leu Glu Val Met Phe Ser  
340 345 350

Arg Ile Val Asn Gln Thr Pro Phe Glu Ile Thr Arg Met Ser Ser Gly  
355 360 365

Gln Met Val Glu Tyr Leu Leu Met Lys Arg Ala Phe Lys Glu Asn Met  
370 375 380

Ile Val Pro Asn Lys Pro Asp Glu Glu Glu Tyr Arg Arg Arg Val Leu  
385 390 395 400

Thr Thr Tyr Glu Gly Gly Tyr Val Lys Glu Pro Glu Lys Gly Met Phe  
405 410 415

Glu Asp Ile Ile Ser Met Asp Phe Arg Cys His Pro Lys Gly Thr Lys  
420 425 430

Val Val Val Lys Gly Lys Gly Ile Val Asn Ile Glu Asp Val Lys Glu  
435 440 445

Gly Asn Tyr Val Leu Gly Ile Asp Gly Trp Gln Lys Val Lys Lys Val  
450 455 460

Trp Lys Tyr Glu Tyr Glu Gly Glu Leu Ile Asn Val Asn Gly Leu Lys  
465 470 475 480

Cys Thr Pro Asn His Lys Ile Pro Leu Arg Tyr Lys Ile Lys His Lys  
485 490 495

Lys Ile Asn Lys Asn Asp Tyr Leu Val Arg Asp Ile Tyr Ala Lys Ser  
500 505 510

Leu Leu Thr Lys Phe Lys Gly Glu Gly Lys Leu Ile Leu Cys Lys Asp  
515 520 525

Phe Glu Thr Ile Gly Asn Tyr Glu Lys Tyr Ile Asn Asp Met Asp Glu  
530 535 540

Asp Phe Ile Leu Lys Ser Glu Leu Ile Gly Ile Leu Leu Ala Glu Gly  
545 550 555 560

His Leu Leu Arg Arg Asp Ile Glu Tyr Phe Asp Ser Ser Arg Gly Lys  
565 570 575

Lys Arg Ile Ser His Gln Tyr Arg Val Glu Ile Thr Val Asn Glu Asp  
580 585 590

# DE 198 40 771 A 1

Glu Lys Asp Phe Ile Glu Lys Ile Lys Tyr Ile Phe Lys Lys Leu Phe  
 595 600 605  
 5 Asn Tyr Glu Leu Tyr Val Arg Arg Lys Lys Gly Thr Lys Ala Ile Thr  
 610 615 620  
 10 Leu Gly Cys Ala Lys Lys Asp Ile Tyr Leu Lys Ile Glu Glu Ile Leu  
 625 630 635 640  
 Lys Asn Lys Glu Lys Tyr Leu Pro Asn Ala Ile Leu Arg Gly Phe Phe  
 15 645 650 655  
 Glu Gly Asp Gly Tyr Val Asn Thr Val Arg Arg Ala Val Val Val Asn  
 20 660 665 670  
 Gln Gly Thr Asn Asn Tyr Asp Lys Ile Lys Phe Ile Ala Ser Leu Leu  
 675 680 685  
 25 Asp Arg Leu Gly Ile Lys Tyr Ser Phe Tyr Thr Tyr Ser Tyr Glu Glu  
 690 695 700  
 Arg Gly Lys Lys Leu Lys Arg Tyr Val Ile Glu Ile Phe Ser Lys Gly  
 30 705 710 715 720  
 Asp Leu Ile Lys Phe Ser Ile Leu Ile Ser Phe Ile Ser Arg Arg Lys  
 35 725 730 735  
 Asn Asn Leu Leu Asn Glu Ile Ile Arg Gln Lys Thr Leu Tyr Lys Ile  
 40 740 745 750  
 Gly Asp Tyr Gly Phe Tyr Asp Leu Asp Asp Val Cys Val Ser Leu Glu  
 755 760 765  
 45 Ser Tyr Lys Gly Glu Val Tyr Asp Leu Thr Leu Glu Gly Arg Pro Tyr  
 770 775 780  
 Tyr Phe Ala Asn Gly Ile Leu Thr His Asn Ser Leu Tyr Pro Ser Ile  
 50 785 790 795 800  
 Ile Ile Ser Tyr Asn Ile Ser Pro Asp Thr Leu Asp Cys Glu Cys Cys  
 55 805 810 815  
 Lys Asp Val Ser Glu Lys Ile Leu Gly His Trp Phe Cys Lys Lys Lys  
 820 825 830  
 60 Glu Gly Leu Ile Pro Lys Thr Leu Arg Asn Leu Ile Glu Arg Arg Ile  
 835 840 845  
 65



# DE 198 40 771 A 1

Asn Ile Lys Arg Arg Met Lys Lys Met Ala Glu Ile Gly Glu Ile Asn  
850 855 860

Glu Glu Tyr Asn Leu Leu Asp Tyr Glu Gln Lys Ser Leu Lys Ile Leu  
865 870 875 880

Ala Asn Ser Ile Leu Pro Asp Glu Tyr Leu Thr Ile Ile Glu Glu Asp  
885 890 895

Gly Ile Lys Val Val Lys Ile Gly Glu Tyr Ile Asp Asp Leu Met Arg  
900 905 910

Lys His Lys Asp Lys Ile Lys Phe Ser Gly Ile Ser Glu Ile Leu Glu  
915 920 925

Thr Lys Asn Leu Lys Thr Phe Ser Phe Asp Lys Ile Thr Lys Lys Cys  
930 935 940

Glu Ile Lys Lys Val Lys Ala Leu Ile Arg His Pro Tyr Phe Gly Lys  
945 950 955 960

Ala Tyr Lys Ile Lys Leu Arg Ser Gly Arg Thr Ile Lys Val Thr Arg  
965 970 975

Gly His Ser Leu Phe Lys Tyr Glu Asn Gly Lys Ile Val Glu Val Lys  
980 985 990

Gly Asp Asp Val Arg Phe Gly Asp Leu Ile Val Val Pro Lys Lys Leu  
995 1000 1005

Thr Cys Val Asp Lys Glu Val Val Ile Asn Ile Pro Lys Arg Leu Ile  
1010 1015 1020

Asn Ala Asp Glu Glu Glu Ile Lys Asp Leu Val Ile Thr Lys His Lys  
1025 1030 1035 1040

Asp Lys Ala Phe Phe Val Lys Leu Lys Lys Thr Leu Glu Asp Ile Glu  
1045 1050 1055

Asn Asn Lys Leu Lys Val Ile Phe Asp Asp Cys Ile Leu Tyr Leu Lys  
1060 1065 1070

Glu Leu Gly Leu Ile Asp Tyr Asn Ile Ile Lys Lys Ile Asn Lys Val  
1075 1080 1085

Asp Ile Lys Ile Leu Asp Glu Glu Lys Phe Lys Ala Tyr Lys Lys Tyr  
1090 1095 1100

# DE 198 40 771 A 1

Phe Asp Thr Val Ile Glu His Gly Asn Phe Lys Lys Gly Arg Cys Asn  
 1105 1110 1115 1120  
 5 Ile Gln Tyr Ile Lys Ile Lys Asp Tyr Ile Ala Asn Ile Pro Asp Lys  
 1125 1130 1135  
 10 Glu Phe Glu Asp Cys Glu Ile Gly Ala Tyr Ser Gly Lys Ile Asn Ala  
 1140 1145 1150  
 15 Leu Leu Lys Leu Asp Glu Lys Leu Ala Lys Phe Leu Gly Phe Phe Val  
 1155 1160 1165  
 20 Thr Arg Gly Arg Leu Lys Lys Gln Lys Leu Lys Gly Glu Thr Val Tyr  
 1170 1175 1180  
 25 Glu Ile Ser Val Tyr Lys Ser Leu Pro Glu Tyr Gln Lys Glu Ile Ala  
 1185 1190 1195 1200  
 30 Glu Thr Phe Lys Glu Val Phe Gly Ala Gly Ser Met Val Lys Asp Lys  
 1205 1210 1215  
 35 Val Thr Met Asp Asn Lys Ile Val Tyr Leu Val Leu Lys Tyr Ile Phe  
 1220 1225 1230  
 40 Lys Cys Gly Asp Lys Asp Lys Lys His Ile Pro Glu Glu Leu Phe Leu  
 1235 1240 1245  
 45 Ala Ser Glu Ser Val Ile Lys Ser Phe Leu Asp Gly Phe Leu Lys Ala  
 1250 1255 1260  
 50 Lys Lys Asn Ser His Lys Gly Thr Ser Thr Phe Met Ala Lys Asp Glu  
 1265 1270 1275 1280  
 55 Lys Tyr Leu Asn Gln Leu Met Ile Leu Phe Asn Leu Val Gly Ile Pro  
 1285 1290 1295  
 60 Thr Arg Phe Thr Pro Val Lys Asn Lys Gly Tyr Lys Leu Thr Leu Asn  
 1300 1305 1310  
 65 Pro Lys Tyr Gly Thr Val Lys Asp Leu Met Leu Asp Glu Val Lys Glu  
 1315 1320 1325  
 70 Ile Glu Ala Phe Glu Tyr Ser Gly Tyr Val Tyr Asp Leu Ser Val Glu  
 1330 1335 1340  
 75 Asp Asn Glu Asn Phe Leu Val Asn Asn Ile Tyr Ala His Asn Ser Val  
 1345 1350 1355 1360

# DE 198 40 771 A 1

Tyr Gly Tyr Leu Ala Phe Pro Arg Ala Arg Phe Tyr Ser Arg Glu Cys  
 1365 1370 1375  
 Ala Glu Ile Val Thr Tyr Leu Gly Arg Lys Tyr Ile Leu Glu Thr Val  
 1380 1385 1390  
 Lys Glu Ala Glu Lys Phe Gly Phe Lys Val Leu Tyr Ile Asp Thr Asp  
 1395 1400 1405  
 Gly Phe Tyr Ala Ile Trp Lys Glu Lys Ile Ser Lys Glu Glu Leu Ile  
 1410 1415 1420  
 Lys Lys Ala Met Glu Phe Val Glu Tyr Ile Asn Ser Lys Leu Pro Gly  
 1425 1430 1435 1440  
 Thr Met Glu Leu Glu Phe Glu Gly Tyr Phe Lys Arg Gly Ile Phe Val  
 1445 1450 1455  
 Thr Lys Lys Arg Tyr Ala Leu Ile Asp Glu Asn Gly Arg Val Thr Val  
 1460 1465 1470  
 Lys Gly Leu Glu Phe Val Arg Arg Asp Trp Ser Asn Ile Ala Lys Ile  
 1475 1480 1485  
 Thr Gln Arg Arg Val Leu Glu Ala Leu Leu Val Glu Gly Ser Ile Glu  
 1490 1495 1500  
 Lys Ala Lys Lys Ile Ile Gln Asp Val Ile Lys Asp Leu Arg Glu Lys  
 1505 1510 1515 1520  
 Lys Ile Lys Lys Glu Asp Leu Ile Ile Tyr Thr Gln Leu Thr Lys Asp  
 1525 1530 1535  
 Pro Lys Glu Tyr Lys Thr Thr Ala Pro His Val Glu Ile Ala Lys Lys  
 1540 1545 1550  
 Leu Met Arg Glu Gly Lys Arg Ile Lys Val Gly Asp Ile Ile Gly Tyr  
 1555 1560 1565  
 Ile Ile Val Lys Gly Thr Lys Ser Ile Ser Glu Arg Ala Lys Leu Pro  
 1570 1575 1580  
 Glu Glu Val Asp Ile Asp Asp Ile Asp Val Asn Tyr Tyr Ile Asp Asn  
 1585 1590 1595 1600  
 Gln Ile Leu Pro Pro Val Leu Arg Ile Met Glu Ala Val Gly Val Ser  
 1605 1610 1615

5  
 10  
 15  
 20  
 25  
 30  
 35  
 40  
 45  
 50  
 55  
 60  
 65

Lys Asn Glu Leu Lys Lys Glu Gly Ala Gln Leu Thr Leu Asp Lys Phe  
 1620 1625 1630

Phe Lys

(2) ANGABEN ZU SEQ ID NO: 25:

(i) SEQUENZKENNZEICHEN:

(A) LÄNGE: 1235 Aminosäuren

(B) ART: Aminosäure

(C) STRANGFORM: Einzelstrang

(D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: Protein

(vi) URSPRÜNGLICHE HERKUNFT:

(A) ORGANISMUS: Pyrococcus horikoshii

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 25:

Met Ile Leu Asp Ala Asp Tyr Ile Thr Glu Asp Gly Lys Pro Ile Ile  
 1 5 10 15

Arg Ile Phe Lys Lys Glu Asn Gly Glu Phe Lys Val Glu Tyr Asp Arg  
 20 25 30

Asn Phe Arg Pro Tyr Ile Tyr Ala Leu Leu Arg Asp Asp Ser Ala Ile  
 35 40 45

Asp Glu Ile Lys Lys Ile Thr Ala Gln Arg His Gly Lys Val Val Arg  
 50 55 60

Ile Val Glu Thr Glu Lys Ile Gln Arg Lys Phe Leu Gly Arg Pro Ile  
 65 70 75 80

Glu Val Trp Lys Leu Tyr Leu Glu His Pro Gln Asp Val Pro Ala Ile  
 85 90 95

Arg Asp Lys Ile Arg Glu His Pro Ala Val Val Asp Ile Phe Glu Tyr  
 100 105 110

Asp Ile Pro Phe Ala Lys Arg Tyr Leu Ile Asp Lys Gly Leu Thr Pro  
 115 120 125

# DE 198 40 771 A 1

Met Glu Gly Asn Glu Lys Leu Thr Phe Leu Ala Val Asp Ile Glu Thr	
130 135 140	
Leu Tyr His Glu Gly Glu Glu Phe Gly Lys Gly Pro Val Ile Met Ile	5
145 150 155 160	
Ser Tyr Ala Asp Glu Glu Gly Ala Lys Val Ile Thr Trp Lys Lys Ile	10
165 170 175	
Asp Leu Pro Tyr Val Glu Val Val Ser Ser Glu Arg Glu Met Ile Lys	15
180 185 190	
Arg Leu Ile Arg Val Ile Lys Glu Lys Asp Pro Asp Val Ile Ile Thr	20
195 200 205	
Tyr Asn Gly Asp Asn Phe Asp Phe Pro Tyr Leu Leu Lys Arg Ala Glu	25
210 215 220	
Lys Leu Gly Ile Lys Leu Leu Leu Gly Arg Asp Asn Ser Glu Pro Lys	30
225 230 235 240	
Met Gln Lys Met Gly Asp Ser Leu Ala Val Glu Ile Lys Gly Arg Ile	35
245 250 255	
His Phe Asp Leu Phe Pro Val Ile Arg Arg Thr Ile Asn Leu Pro Thr	40
260 265 270	
Tyr Thr Leu Glu Ala Val Tyr Glu Ala Ile Phe Gly Lys Pro Lys Glu	45
275 280 285	
Lys Val Tyr Ala Asp Glu Ile Ala Lys Ala Trp Glu Thr Gly Glu Gly	50
290 295 300	
Leu Glu Arg Val Ala Lys Tyr Ser Met Glu Asp Ala Lys Val Thr Tyr	55
305 310 315 320	
Glu Leu Gly Arg Glu Phe Phe Pro Met Glu Ala Gln Leu Ala Arg Leu	60
325 330 335	
Val Gly Gln Pro Val Trp Asp Val Ser Arg Ser Ser Thr Gly Asn Leu	65
340 345 350	
Val Glu Trp Phe Leu Leu Arg Lys Ala Tyr Glu Arg Asn Glu Leu Ala	
355 360 365	
Pro Asn Lys Pro Asp Glu Lys Glu Tyr Glu Arg Arg Leu Arg Glu Ser	
370 375 380	

DE 198 40 771 A 1

Tyr Glu Gly Gly Tyr Val Lys Glu Pro Glu Lys Gly Leu Trp Glu Gly  
 385 390 395 400  
 5 Ile Val Ser Leu Asp Phe Arg Ser Leu Tyr Pro Ser Ile Ile Ile Thr  
 405 410 415  
 10 His Asn Val Ser Pro Asp Thr Leu Asn Arg Glu Gly Cys Glu Glu Tyr  
 420 425 430  
 15 Asp Val Ala Pro Lys Val Gly His Arg Phe Cys Lys Asp Phe Pro Gly  
 435 440 445  
 20 Phe Ile Pro Ser Leu Leu Gly Gln Leu Leu Glu Glu Arg Gln Lys Ile  
 450 455 460  
 25 Lys Lys Arg Met Lys Glu Ser Lys Asp Pro Val Glu Lys Lys Leu Leu  
 465 470 475 480  
 30 Asp Tyr Arg Gln Arg Ala Ile Lys Ile Leu Ala Asn Ser Ile Leu Pro  
 485 490 495  
 35 Asp Glu Trp Leu Pro Ile Val Glu Asn Glu Lys Val Arg Phe Val Lys  
 500 505 510  
 40 Ile Gly Asp Phe Ile Asp Arg Glu Ile Glu Glu Asn Ala Glu Arg Val  
 515 520 525  
 45 Lys Arg Asp Gly Glu Thr Glu Ile Leu Glu Val Lys Asp Leu Lys Ala  
 530 535 540  
 50 Leu Ser Phe Asn Arg Glu Thr Lys Lys Ser Glu Leu Lys Lys Val Lys  
 545 550 555 560  
 55 Ala Leu Ile Arg His Arg Tyr Ser Gly Lys Val Tyr Ser Ile Lys Leu  
 565 570 575  
 60 Lys Ser Gly Arg Arg Ile Lys Ile Thr Ser Gly His Ser Leu Phe Ser  
 580 585 590  
 65 Val Lys Asn Gly Lys Leu Val Lys Val Arg Gly Asp Glu Leu Lys Pro  
 595 600 605  
 70 Gly Asp Leu Val Val Val Pro Gly Arg Leu Lys Leu Pro Glu Ser Lys  
 610 615 620  
 75 Gln Val Leu Asn Leu Val Glu Leu Leu Leu Lys Leu Pro Glu Glu Glu  
 625 630 635 640

# DE 198 40 771 A 1

Thr Ser Asn Ile Val Met Met Ile Pro Val Lys Gly Arg Lys Asn Phe	
645 650 655	
Phe Lys Gly Met Leu Lys Thr Leu Tyr Trp Ile Phe Gly Glu Gly Glu	5
660 665 670	
Arg Pro Arg Thr Ala Gly Arg Tyr Leu Lys His Leu Glu Arg Leu Gly	10
675 680 685	
Tyr Val Lys Leu Lys Arg Arg Gly Cys Glu Val Leu Asp Trp Glu Ser	15
690 695 700	
Leu Lys Arg Tyr Arg Lys Leu Tyr Glu Thr Leu Ile Lys Asn Leu Lys	20
705 710 715 720	
Tyr Asn Gly Asn Ser Arg Ala Tyr Met Val Glu Phe Asn Ser Leu Arg	25
725 730 735	
Asp Val Val Ser Leu Met Pro Ile Glu Glu Leu Lys Glu Trp Ile Ile	30
740 745 750	
Gly Glu Pro Arg Gly Pro Lys Ile Gly Thr Phe Ile Asp Val Asp Asp	35
755 760 765	
Ser Phe Ala Lys Leu Leu Gly Tyr Tyr Ile Ser Ser Gly Asp Val Glu	40
770 775 780	
Lys Asp Arg Val Lys Phe His Ser Lys Asp Gln Asn Val Leu Glu Asp	45
785 790 795 800	
Ile Ala Lys Leu Ala Glu Lys Leu Phe Gly Lys Val Arg Arg Gly Arg	50
805 810 815	
Gly Tyr Ile Glu Val Ser Gly Lys Ile Ser His Ala Ile Phe Arg Val	55
820 825 830	
Leu Ala Glu Gly Lys Arg Ile Pro Glu Phe Ile Phe Thr Ser Pro Met	60
835 840 845	
Asp Ile Lys Val Ala Phe Leu Lys Gly Leu Asn Gly Asn Ala Glu Glu	65
850 855 860	
Leu Thr Phe Ser Thr Lys Ser Glu Leu Leu Val Asn Gln Leu Ile Leu	
865 870 875 880	
Leu Leu Asn Ser Ile Gly Val Ser Asp Ile Lys Ile Glu His Glu Lys	
885 890 895	

# DE 198 40 771 A 1

Gly Val Tyr Arg Val Tyr Ile Asn Lys Lys Glu Ser Ser Asn Gly Asp  
 900 905 910

5 Ile Val Leu Asp Ser Val Glu Ser Ile Glu Val Glu Lys Tyr Glu Gly  
 915 920 925

10 Tyr Val Tyr Asp Leu Ser Val Glu Asp Asn Glu Asn Phe Leu Val Gly  
 930 935 940

Phe Gly Leu Leu Tyr Ala His Asn Ser Tyr Tyr Gly Tyr Tyr Gly Tyr  
 15 945 950 955 960

Ala Lys Ala Arg Trp Tyr Cys Lys Glu Cys Ala Glu Ser Val Thr Ala  
 20 965 970 975

Trp Gly Arg Gln Tyr Ile Asp Leu Val Arg Arg Glu Leu Glu Ala Arg  
 25 980 985 990

Gly Phe Lys Val Leu Tyr Ile Asp Thr Asp Gly Leu Tyr Ala Thr Ile  
 995 1000 1005

30 Pro Gly Val Lys Asp Trp Glu Glu Val Lys Arg Arg Ala Leu Glu Phe  
 1010 1015 1020

Val Asp Tyr Ile Asn Ser Lys Leu Pro Gly Val Leu Glu Leu Glu Tyr  
 35 1025 1030 1035 1040

Glu Gly Phe Tyr Ala Arg Gly Phe Phe Val Thr Lys Lys Lys Tyr Ala  
 40 1045 1050 1055

Leu Ile Asp Glu Glu Gly Lys Ile Val Thr Arg Gly Leu Glu Ile Val  
 1060 1065 1070

45 Arg Arg Asp Trp Ser Glu Ile Ala Lys Glu Thr Gln Ala Arg Val Leu  
 1075 1080 1085

Glu Ala Ile Leu Lys His Gly Asn Val Glu Glu Ala Val Lys Ile Val  
 50 1090 1095 1100

Lys Asp Val Thr Glu Lys Leu Thr Asn Tyr Glu Val Pro Pro Glu Lys  
 55 1105 1110 1115 1120

Leu Val Ile Tyr Glu Gln Ile Thr Arg Pro Ile Asn Glu Tyr Lys Ala  
 1125 1130 1135

60 Ile Gly Pro His Val Ala Val Ala Lys Arg Leu Met Ala Arg Gly Ile  
 1140 1145 1150

65



# DE 198 40 771 A 1

Lys Val Lys Pro Gly Met Val Ile Gly Tyr Ile Val Leu Arg Gly Asp  
1155 1160 1165

Gly Pro Ile Ser Lys Arg Ala Ile Ser Ile Glu Glu Phe Asp Pro Arg  
1170 1175 1180

Lys His Lys Tyr Asp Ala Glu Tyr Tyr Ile Glu Asn Gln Val Leu Pro  
1185 1190 1195 1200

Ala Val Glu Arg Ile Leu Lys Ala Phe Gly Tyr Lys Arg Glu Asp Leu  
1205 1210 1215

Arg Trp Gln Lys Thr Lys Gln Val Gly Leu Gly Ala Trp Ile Lys Val  
1220 1225 1230

Lys Lys Ser  
1235

(2) ANGABEN ZU SEQ ID NO: 26:

(i) SEQUENZKENNZEICHEN:

(A) LÄNGE: 586 Aminosäuren

(B) ART: Aminosäure

(C) STRANGFORM: Einzelstrang

(D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: Protein

(vi) URSPRÜNGLICHE HERKUNFT:

(A) ORGANISMUS: *Methanobacterium thermoautotrophicum*

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 26:

Met Glu Asp Tyr Arg Met Val Leu Leu Asp Ile Asp Tyr Val Thr Val  
1 5 10 15

Asp Glu Val Pro Val Ile Arg Leu Phe Gly Lys Asp Lys Ser Gly Gly  
20 25 30

Asn Glu Pro Ile Ile Ala His Asp Arg Ser Phe Arg Pro Tyr Ile Tyr  
35 40 45

Ala Ile Pro Thr Asp Leu Asp Glu Cys Leu Arg Glu Leu Glu Glu Leu  
50 55 60

# DE 198 40 771 A 1

Glu Leu Glu Lys Leu Glu Val Lys Glu Met Arg Asp Leu Gly Arg Pro  
65 70 75 80

5 Thr Glu Val Ile Arg Ile Glu Phe Arg His Pro Gln Asp Val Pro Lys  
85 90 95

10 Ile Arg Asp Arg Ile Arg Asp Leu Glu Ser Val Arg Asp Ile Arg Glu  
100 105 110

15 His Asp Ile Pro Phe Tyr Arg Arg Tyr Leu Ile Asp Lys Ser Ile Val  
115 120 125

20 Pro Met Glu Glu Leu Glu Phe Gln Gly Val Glu Val Asp Ser Ala Pro  
130 135 140

Ser Val Thr Thr Asp Val Arg Thr Val Glu Val Thr Gly Arg Val Gln  
145 150 155 160

25 Ser Thr Gly Ser Gly Ala His Gly Leu Asp Ile Leu Ser Phe Asp Ile  
165 170 175

30 Glu Val Arg Asn Pro His Gly Met Pro Asp Pro Glu Lys Asp Glu Ile  
180 185 190

35 Val Met Ile Gly Val Ala Gly Asn Met Gly Tyr Glu Ser Val Ile Ser  
195 200 205

Thr Ala Gly Asp His Leu Asp Phe Val Glu Val Val Glu Asp Glu Arg  
210 215 220

40 Glu Leu Leu Glu Arg Phe Ala Glu Ile Val Ile Asp Lys Lys Pro Asp  
225 230 235 240

45 Ile Leu Val Gly Tyr Asn Ser Asp Asn Phe Asp Phe Pro Tyr Ile Thr  
245 250 255

50 Arg Arg Ala Ala Ile Leu Gly Ala Glu Leu Asp Leu Gly Trp Asp Gly  
260 265 270

55 Ser Lys Ile Arg Thr Met Arg Arg Gly Phe Ala Asn Ala Thr Ala Ile  
275 280 285

Lys Gly Thr Val His Val Asp Leu Tyr Pro Val Met Arg Arg Tyr Met  
290 295 300

60 Asn Leu Asp Arg Tyr Thr Leu Glu Arg Val Tyr Gln Glu Leu Phe Gly  
305 310 315 320

65

# DE 198 40 771 A 1

Glu Glu Lys Ile Asp Leu Pro Gly Asp Arg Leu Trp Glu Tyr Trp Asp  
325 330 335

Arg Asp Glu Leu Arg Asp Glu Leu Phe Arg Tyr Ser Leu Asp Asp Val  
340 345 350

Val Ala Thr His Arg Ile Ala Glu Lys Ile Leu Pro Leu Asn Leu Glu  
355 360 365

Leu Thr Arg Leu Val Gly Gln Pro Leu Phe Asp Ile Ser Arg Met Ala  
370 375 380

Thr Gly Gln Gln Ala Glu Trp Phe Leu Val Arg Lys Ala Tyr Gln Tyr  
385 390 395 400

Gly Glu Leu Val Pro Asn Lys Pro Ser Gln Ser Asp Phe Ser Ser Arg  
405 410 415

Arg Gly Arg Arg Ala Val Gly Gly Tyr Val Lys Glu Pro Glu Lys Gly  
420 425 430

Leu His Glu Asn Ile Val Gln Phe Asp Phe Arg Ser Leu Tyr Pro Ser  
435 440 445

Ile Ile Ile Ser Lys Asn Ile Ser Pro Asp Thr Leu Thr Asp Asp Glu  
450 455 460

Glu Ser Glu Cys Tyr Val Ala Pro Glu Tyr Gly Tyr Arg Phe Arg Lys  
465 470 475 480

Ser Pro Arg Gly Phe Val Pro Ser Val Ile Gly Glu Ile Leu Ser Glu  
485 490 495

Arg Val Arg Ile Lys Glu Glu Met Lys Gly Ser Asp Asp Pro Met Glu  
500 505 510

Arg Lys Ile Leu Asn Val Gln Gln Glu Ala Leu Lys Arg Leu Ala Asn  
515 520 525

Thr Met Tyr Gly Val Tyr Gly Tyr Ser Arg Phe Arg Trp Tyr Ser Met  
530 535 540

Glu Cys Ala Glu Ala Ile Thr Ala Trp Gly Arg Asp Tyr Ile Lys Lys  
545 550 555 560

Thr Ile Lys Thr Ala Glu Glu Phe Gly Phe His Thr Val Tyr Ala Asp  
565 570 575

Thr Asp Gly Phe Tyr Ala Thr Tyr Arg Gly  
580 585

5 (2) ANGABEN ZU SEQ ID NO: 27:

(i) SEQUENZKENNZEICHEN:

- 10 (A) LÄNGE: 1143 Aminosäuren  
(B) ART: Aminosäure  
(C) STRANGFORM: Einzelstrang  
(D) TOPOLOGIE: linear

15 (ii) ART DES MOLEKÜLS: Protein

(vi) URSPRÜNGLICHE HERKUNFT:

- 20 (A) ORGANISMUS: *Archaeoglobus fulgidus*

25 (xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 27:

Met Asp Ala Thr Leu Asp Arg Phe Phe Pro Leu Phe Glu Ser Glu Ser  
1 5 10 15

Asn Glu Asp Phe Trp Arg Ile Glu Glu Ile Arg Arg Tyr His Glu Ser  
20 25 30

Leu Met Val Glu Leu Asp Arg Ile Tyr Arg Ile Ala Glu Ala Ala Arg  
35 40 45

Lys Lys Gly Leu Asp Pro Glu Leu Ser Val Glu Ile Pro Ile Ala Lys  
50 55 60

Asn Met Ala Glu Arg Val Glu Lys Leu Met Asn Leu Gln Gly Leu Ala  
65 70 75 80

Lys Arg Ile Met Glu Leu Glu Glu Gly Gly Leu Ser Arg Glu Leu Ile  
85 90 95

Cys Phe Lys Val Ala Asp Glu Ile Val Glu Gly Lys Phe Gly Glu Met  
100 105 110

Pro Lys Glu Glu Ala Ile Asp Lys Ala Val Arg Thr Ala Val Ala Ile  
115 120 125

Met Thr Glu Gly Val Val Ala Ala Pro Ile Glu Gly Ile Ala Arg Val  
130 135 140

# DE 198 40 771 A 1

Arg Ile Asp Arg Glu Asn Phe Leu Arg Val Tyr Tyr Ala Gly Pro Ile  
145 150 155 160

Arg Ser Ala Gly Gly Thr Ala Gln Val Ile Ser Val Leu Val Ala Asp  
165 170 175

Tyr Val Arg Arg Lys Ala Glu Ile Gly Arg Tyr Val Pro Thr Glu Glu  
180 185 190

Glu Ile Leu Arg Tyr Cys Glu Glu Ile Pro Leu Tyr Lys Lys Val Ala  
195 200 205

Asn Leu Gln Tyr Leu Pro Ser Asp Glu Glu Ile Arg Leu Ile Val Ser  
210 215 220

Asn Cys Pro Ile Cys Ile Asp Gly Glu Pro Thr Glu Ser Ala Glu Val  
225 230 235 240

Ser Gly Tyr Arg Asn Leu Pro Arg Val Glu Thr Asn Arg Val Arg Gly  
245 250 255

Gly Met Ala Leu Val Ile Ala Glu Gly Ile Ala Leu Lys Ala Pro Lys  
260 265 270

Leu Lys Lys Met Val Asp Glu Val Gly Ile Glu Gly Trp Glu Trp Leu  
275 280 285

Asp Ala Leu Ile Lys Gly Gly Gly Asp Ser Gly Ser Glu Glu Glu Lys  
290 295 300

Ala Val Ile Lys Pro Lys Asp Lys Tyr Leu Ser Asp Ile Val Ala Gly  
305 310 315 320

Arg Pro Val Leu Ser His Pro Ser Arg Lys Gly Gly Phe Arg Leu Arg  
325 330 335

Tyr Gly Arg Ala Arg Asn Ser Gly Phe Ala Thr Val Gly Val Asn Pro  
340 345 350

Ala Thr Met Tyr Leu Leu Glu Phe Val Ala Val Gly Thr Gln Leu Lys  
355 360 365

Val Glu Arg Pro Gly Lys Ala Gly Gly Val Val Pro Val Ser Thr Ile  
370 375 380

Glu Gly Pro Thr Val Arg Leu Lys Asn Gly Asp Val Val Lys Ile Asn  
385 390 395 400

5

10

15

20

25

30

35

40

45

50

55

60

65

# DE 198 40 771 A 1

Thr Leu Ser Glu Ala Lys Ala Leu Lys Gly Glu Val Ala Ala Ile Leu  
 405 410 415  
 5 Asp Leu Gly Glu Ile Leu Ile Asn Tyr Gly Asp Phe Leu Glu Asn Asn  
 420 425 430  
 10 His Pro Leu Ile Pro Ala Ser Tyr Thr Tyr Glu Trp Trp Ile Gln Glu  
 435 440 445  
 Ala Glu Lys Ala Gly Leu Arg Gly Asp Tyr Arg Lys Ile Ser Glu Glu  
 15 450 455 460  
 Glu Ala Leu Lys Leu Cys Asp Glu Phe His Val Pro Leu His Pro Asp  
 20 465 470 475 480  
 Tyr Thr Tyr Leu Trp His Asp Ile Ser Val Glu Asp Tyr Arg Tyr Leu  
 485 490 495  
 25 Arg Asn Phe Val Ser Asp Asn Gly Lys Ile Glu Gly Lys His Gly Lys  
 500 505 510  
 Ser Val Leu Leu Leu Pro Tyr Asp Ser Arg Val Lys Glu Ile Leu Glu  
 30 515 520 525  
 Ala Leu Leu Leu Glu His Lys Val Arg Glu Ser Phe Ile Val Ile Glu  
 35 530 535 540  
 Thr Trp Arg Ala Phe Ile Arg Cys Leu Gly Leu Asp Glu Lys Leu Ser  
 545 550 555 560  
 40 Lys Val Ser Glu Val Ser Gly Lys Asp Val Leu Glu Ile Val Asn Gly  
 565 570 575  
 Ile Ser Gly Ile Lys Val Arg Pro Lys Ala Leu Ser Arg Ile Gly Ala  
 45 580 585 590  
 Arg Met Gly Arg Pro Glu Lys Ala Lys Glu Arg Lys Met Ser Pro Pro  
 50 595 600 605  
 Pro His Ile Leu Phe Pro Val Gly Met Ala Gly Gly Asn Thr Arg Asp  
 55 610 615 620  
 Ile Lys Asn Ala Ile Asn Tyr Thr Lys Ser Tyr Asn Ala Lys Lys Gly  
 625 630 635 640  
 60 Glu Ile Glu Val Glu Ile Ala Ile Arg Lys Cys Pro Gln Cys Gly Lys  
 645 650 655  
 65

# DE 198 40 771 A 1

Glu Thr Phe Trp Leu Lys Cys Asp Val Cys Gly Glu Leu Thr Glu Gln  
660 665 670

Leu Tyr Tyr Cys Pro Ser Cys Arg Met Lys Asn Thr Ser Ser Val Cys  
675 680 685

Glu Ser Cys Gly Arg Glu Cys Glu Gly Tyr Met Lys Arg Lys Val Asp  
690 695 700

Leu Arg Glu Leu Tyr Glu Glu Ala Ile Ala Asn Leu Gly Glu Tyr Asp  
705 710 715 720

Ser Phe Asp Thr Ile Lys Gly Val Lys Gly Met Thr Ser Lys Thr Lys  
725 730 735

Ile Pro Glu Arg Leu Glu Lys Gly Ile Leu Arg Val Lys His Gly Val  
740 745 750

Phe Val Phe Lys Asp Gly Thr Ala Arg Phe Asp Ala Thr Asp Leu Pro  
755 760 765

Ile Thr His Phe Lys Pro Ala Glu Ile Gly Val Ser Val Glu Lys Leu  
770 775 780

Arg Glu Leu Gly Tyr Glu Arg Asp Tyr Lys Gly Ala Glu Leu Lys Asn  
785 790 795 800

Glu Asn Gln Ile Val Glu Leu Lys Pro Gln Asp Val Ile Leu Pro Lys  
805 810 815

Ser Gly Ala Glu Tyr Leu Leu Arg Val Ala Asn Phe Ile Asp Asp Leu  
820 825 830

Leu Val Lys Phe Tyr Lys Met Glu Pro Phe Tyr Asn Ala Lys Ser Val  
835 840 845

Glu Asp Leu Ile Gly His Leu Val Ile Gly Leu Ala Pro His Thr Ser  
850 855 860

Ala Gly Val Leu Gly Arg Ile Ile Gly Phe Ser Asp Val Leu Ala Gly  
865 870 875 880

Tyr Ala His Pro Tyr Phe His Ala Ala Lys Arg Arg Asn Cys Asp Gly  
885 890 895

Asp Glu Asp Cys Phe Met Leu Leu Leu Asp Gly Leu Leu Asn Phe Ser  
900 905 910

# DE 198 40 771 A 1

Arg Lys Phe Leu Pro Asp Lys Arg Gly Gly Gln Met Asp Ala Pro Leu  
915 920 925

Val Leu Thr Ala Ile Val Asp Pro Arg Glu Val Asp Lys Glu Val His  
930 935 940

Asn Met Asp Ile Val Glu Arg Tyr Pro Leu Glu Phe Tyr Glu Ala Thr  
945 950 955 960

Met Arg Phe Ala Ser Pro Lys Glu Met Glu Asp Tyr Val Glu Lys Val  
965 970 975

Lys Asp Arg Leu Lys Asp Glu Ser Arg Phe Cys Gly Leu Phe Phe Thr  
980 985 990

His Asp Thr Glu Asn Ile Ala Ala Gly Val Lys Glu Ser Ala Tyr Lys  
995 1000 1005

Ser Leu Lys Thr Met Gln Asp Lys Val Tyr Arg Gln Met Glu Leu Ala  
1010 1015 1020

Arg Met Ile Val Ala Val Asp Glu His Asp Val Ala Glu Arg Val Ile  
1025 1030 1035 1040

Asn Val His Phe Leu Pro Asp Ile Ile Gly Asn Leu Arg Ala Phe Ser  
1045 1050 1055

Arg Gln Glu Phe Arg Cys Thr Arg Cys Asn Thr Lys Tyr Arg Arg Ile  
1060 1065 1070

Pro Leu Val Gly Lys Cys Leu Lys Cys Gly Asn Lys Leu Thr Leu Thr  
1075 1080 1085

Val His Ser Ser Ser Ile Met Lys Tyr Leu Glu Leu Ser Lys Phe Leu  
1090 1095 1100

Cys Glu Asn Phe Asn Val Ser Ser Tyr Thr Lys Gln Arg Leu Met Leu  
1105 1110 1115 1120

Leu Glu Gln Glu Ile Lys Ser Met Phe Glu Asn Gly Thr Glu Lys Gln  
1125 1130 1135

Val Ser Ile Ser Asp Phe Val  
1140

(2) ANGABEN ZU SEQ ID NO: 28:

(i) SEQUENZKENNZEICHEN:



(A) LÄNGE: 1139 Aminosäuren  
(B) ART: Aminosäure  
(C) STRANGFORM: Einzelstrang  
(D) TOPOLOGIE: linear

5

(ii) ART DES MOLEKÜLS: Protein

(vi) URSPRÜNGLICHE HERKUNFT:

10

(A) ORGANISMUS: Methanococcus jannaschii

15

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 28:

Met Ile Val Met Val His Val Ala Cys Ser Glu Asn Met Lys Lys Tyr  
1 5 10 15

20

Phe Glu Asn Ile Val Asp Glu Val Lys Lys Ile Tyr Arg Ile Ala Glu  
20 25 30

25

Glu Cys Arg Lys Lys Gly Phe Asp Pro Thr Asp Glu Val Glu Ile Pro  
35 40 45

30

Leu Ala Ala Asp Met Ala Asp Arg Val Glu Gly Leu Val Gly Pro Lys  
50 55 60

35

Gly Val Ala Glu Arg Ile Arg Glu Leu Val Lys Glu Leu Gly Lys Glu  
65 70 75 80

Pro Ala Ala Leu Glu Ile Ala Lys Glu Ile Val Glu Gly Lys Phe Gly  
85 90 95

40

Asn Phe Asp Lys Glu Lys Lys Ala Glu Gln Ala Val Arg Thr Ala Leu  
100 105 110

45

Ala Val Leu Thr Glu Gly Ile Val Ala Ala Pro Leu Glu Gly Ile Ala  
115 120 125

50

Asp Val Lys Ile Lys Lys Asn Pro Asp Gly Thr Glu Tyr Leu Ala Ile  
130 135 140

Tyr Tyr Ala Gly Pro Ile Arg Ser Ala Gly Gly Thr Ala Gln Ala Leu  
145 150 155 160

55

Ser Val Leu Val Gly Asp Phe Val Arg Lys Ala Met Gly Leu Asp Arg  
165 170 175

60

65

# DE 198 40 771 A 1

Tyr Lys Pro Thr Glu Asp Glu Ile Glu Arg Tyr Val Glu Glu Val Glu  
 180 185 190

5 Leu Tyr Gln Ser Glu Val Gly Ser Phe Gln Tyr Asn Pro Thr Ala Asp  
 195 200 205

10 Glu Ile Arg Thr Ala Ile Arg Asn Ile Pro Ile Glu Ile Thr Gly Glu  
 210 215 220

15 Ala Thr Asp Asp Val Glu Val Ser Gly His Arg Asp Leu Pro Arg Val  
 225 230 235 240

20 Glu Thr Asn Gln Leu Arg Gly Gly Ala Leu Leu Val Leu Val Glu Gly  
 245 250 255

25 Val Leu Leu Lys Ala Pro Lys Ile Leu Arg His Val Asp Lys Leu Gly  
 260 265 270

30 Ile Glu Gly Trp Asp Trp Leu Lys Asp Leu Met Ser Lys Lys Glu Glu  
 275 280 285

35 Lys Glu Glu Glu Lys Asp Glu Lys Val Asp Asp Glu Glu Ile Asp Glu  
 290 295 300

40 Glu Glu Glu Glu Ile Ser Gly Tyr Trp Arg Asp Val Lys Ile Glu Ala  
 305 310 315 320

45 Asn Lys Lys Phe Ile Ser Glu Val Ile Ala Gly Arg Pro Val Phe Ala  
 325 330 335

50 His Pro Ser Lys Val Gly Gly Phe Arg Leu Arg Tyr Gly Arg Ser Arg  
 340 345 350

55 Asn Thr Gly Phe Ala Thr Gln Gly Phe His Pro Ala Leu Met Tyr Leu  
 355 360 365

60 Val Asp Glu Phe Met Ala Val Gly Thr Gln Leu Lys Thr Glu Arg Pro  
 370 375 380

65 Gly Lys Ala Thr Cys Val Val Pro Val Asp Ser Ile Glu Pro Pro Ile  
 385 390 395 400

70 Val Lys Leu Lys Asn Gly Asp Val Ile Arg Val Asp Thr Ile Glu Lys  
 405 410 415

75 Ala Met Asp Val Arg Asn Arg Val Glu Glu Ile Leu Phe Leu Gly Asp  
 420 425 430

# DE 198 40 771 A 1

Val Leu Val Asn Tyr Gly Asp Phe Leu Glu Asn Asn His Pro Leu Leu  
435 440 445

Pro Ser Cys Trp Cys Glu Glu Trp Tyr Glu Lys Ile Leu Ile Ala Asn  
450 455 460

Asn Ile Glu Tyr Asp Lys Asp Phe Ile Lys Asn Pro Lys Pro Glu Glu  
465 470 475 480

Ala Val Lys Phe Ala Leu Glu Thr Lys Thr Pro Leu His Pro Arg Phe  
485 490 495

Thr Tyr His Trp His Asp Val Ser Lys Glu Asp Ile Ile Leu Leu Arg  
500 505 510

Asn Trp Leu Leu Lys Gly Lys Glu Asp Ser Leu Glu Gly Lys Lys Val  
515 520 525

Trp Ile Val Asp Leu Glu Ile Glu Glu Asp Lys Lys Ala Lys Arg Ile  
530 535 540

Leu Glu Leu Ile Gly Cys Cys His Leu Val Arg Asn Lys Lys Val Ile  
545 550 555 560

Ile Glu Glu Tyr Tyr Pro Leu Leu Tyr Ser Leu Gly Phe Asp Val Glu  
565 570 575

Asn Lys Lys Asp Leu Val Glu Asn Ile Glu Lys Ile Leu Glu Ser Ala  
580 585 590

Lys Asn Ser Met His Leu Ile Asn Leu Leu Ala Pro Phe Glu Val Arg  
595 600 605

Arg Asn Thr Tyr Val Tyr Val Gly Ala Arg Met Gly Arg Pro Glu Lys  
610 615 620

Ala Ala Pro Arg Lys Met Lys Pro Pro Val Asn Gly Leu Phe Pro Ile  
625 630 635 640

Gly Asn Ala Gly Gly Gln Val Arg Leu Ile Asn Lys Ala Val Glu Glu  
645 650 655

Asn Asn Thr Asp Asp Val Asp Val Ser Tyr Thr Arg Cys Pro Asn Cys  
660 665 670

Gly Lys Ile Ser Leu Tyr Arg Val Cys Pro Phe Cys Gly Thr Lys Val  
675 680 685

5

10

15

20

25

30

35

40

45

50

55

60

65

# DE 198 40 771 A 1

Glu Leu Asp Asn Phe Gly Arg Ile Lys Ala Pro Leu Lys Asp Tyr Trp  
 690 695 700

5 Tyr Ala Ala Leu Lys Arg Leu Gly Ile Asn Lys Pro Gly Asp Val Lys  
 705 710 715 720

10 Cys Ile Lys Gly Met Thr Ser Lys Gln Lys Ile Val Glu Pro Leu Glu  
 725 730 735

15 Lys Ala Ile Leu Arg Ala Ile Asn Glu Val Tyr Val Phe Lys Asp Gly  
 740 745 750

20 Thr Thr Arg Phe Asp Cys Thr Asp Val Pro Val Thr His Phe Lys Pro  
 755 760 765

25 Asn Glu Ile Asn Val Thr Val Glu Lys Leu Arg Glu Leu Gly Tyr Asp  
 770 775 780

30 Lys Asp Ile Tyr Gly Asn Glu Leu Val Asp Gly Glu Gln Val Val Glu  
 785 790 795 800

35 Leu Lys Pro Gln Asp Val Ile Ile Pro Glu Ser Cys Ala Glu Tyr Phe  
 805 810 815

40 Val Lys Val Ala Asn Phe Ile Asp Asp Leu Leu Glu Lys Phe Tyr Lys  
 820 825 830

45 Val Glu Arg Phe Tyr Asn Val Lys Lys Lys Glu Asp Leu Ile Gly His  
 835 840 845

50 Leu Val Ile Gly Met Ala Pro His Thr Ser Ala Gly Met Val Gly Arg  
 850 855 860

55 Ile Ile Gly Tyr Thr Lys Ala Asn Val Gly Tyr Ala His Pro Tyr Phe  
 865 870 875 880

60 His Ala Ala Lys Arg Arg Asn Cys Asp Gly Asp Glu Asp Ser Phe Phe  
 885 890 895

65 Leu Leu Leu Asp Ala Phe Leu Asn Phe Ser Lys Lys Phe Leu Pro Asp  
 900 905 910

70 Lys Arg Gly Gly Gln Met Asp Ala Pro Leu Val Leu Thr Thr Ile Leu  
 915 920 925

75 Asp Pro Lys Glu Val Asp Gly Glu Val His Asn Met Asp Thr Met Trp  
 930 935 940

# DE 198 40 771 A 1

Ser Tyr Pro Leu Glu Phe Tyr Glu Lys Thr Leu Glu Met Pro Ser Pro  
945 950 955 960

Lys Glu Val Lys Glu Phe Met Glu Thr Val Glu Asp Arg Leu Gly Lys  
965 970 975

Pro Glu Gln Tyr Glu Gly Ile Gly Tyr Thr His Glu Thr Ser Arg Ile  
980 985 990

Asp Leu Gly Pro Lys Val Cys Ala Tyr Lys Thr Leu Gly Ser Met Leu  
995 1000 1005

Glu Lys Thr Thr Ser Gln Leu Ser Val Ala Lys Lys Ile Arg Ala Thr  
1010 1015 1020

Asp Glu Arg Asp Val Ala Glu Lys Val Ile Gln Ser His Phe Ile Pro  
1025 1030 1035 1040

Asp Leu Ile Gly Asn Leu Arg Ala Phe Ser Arg Gln Ala Val Arg Cys  
1045 1050 1055

Lys Cys Gly Ala Lys Tyr Arg Arg Ile Pro Leu Lys Gly Lys Cys Pro  
1060 1065 1070

Lys Cys Gly Ser Asn Leu Ile Leu Thr Val Ser Lys Gly Ala Val Glu  
1075 1080 1085

Lys Tyr Met Asp Val Ala Glu Lys Met Ala Glu Glu Tyr Asn Val Asn  
1090 1095 1100

Asp Tyr Ile Lys Gln Arg Leu Lys Ile Ile Lys Glu Gly Ile Asn Ser  
1105 1110 1115 1120

Ile Phe Glu Asn Glu Lys Ser Arg Gln Val Lys Leu Ser Asp Phe Phe  
1125 1130 1135

Lys Ile Gly

## (2) ANGABEN ZU SEQ ID NO: 29:

### (i) SEQUENZKENNZEICHEN:

- (A) LÄNGE: 1434 Aminosäuren
- (B) ART: Aminosäure
- (C) STRANGFORM: Einzelstrang
- (D) TOPOLOGIE: linear

### (ii) ART DES MOLEKÜLS: Protein

## (vi) URSPRÜNGLICHE HERKUNFT:

(A) ORGANISMUS: *Pyrococcus horikoshii*

5

## (xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 29:

10

Met Val Leu Met Glu Leu Pro Lys Glu Met Glu Glu Tyr Phe Ser Met  
 1 5 10 15

15

Leu Gln Arg Glu Ile Asp Lys Ala Tyr Glu Ile Ala Lys Lys Ala Arg  
 20 25 30

20

Ala Gln Gly Lys Asp Pro Ser Leu Asp Val Glu Ile Pro Gln Ala Ser  
 35 40 45

25

Asp Met Ala Gly Arg Val Glu Ser Leu Val Gly Pro Pro Gly Val Ala  
 50 55 60

30

Glu Arg Ile Arg Glu Leu Val Lys Glu Tyr Gly Lys Glu Ile Ala Ala  
 65 70 75 80

Leu Lys Ile Val Asp Glu Ile Ile Asp Gly Lys Phe Gly Asp Leu Gly  
 85 90 95

35

Ser Lys Glu Lys Tyr Ala Glu Gln Ala Val Arg Thr Ala Leu Ala Ile  
 100 105 110

40

Leu Thr Glu Gly Val Val Ser Ala Pro Ile Glu Gly Ile Ala Ser Val  
 115 120 125

45

Lys Ile Lys Arg Asn Thr Trp Ser Asp Asn Ser Glu Tyr Leu Ala Leu  
 130 135 140

Tyr Tyr Ala Gly Pro Ile Arg Ser Ser Gly Gly Thr Ala Gln Ala Leu  
 145 150 155 160

50

Ser Val Leu Val Gly Asp Tyr Val Arg Arg Lys Leu Gly Leu Asp Arg  
 165 170 175

55

Phe Lys Pro Ser Glu Lys His Ile Glu Arg Met Val Glu Glu Val Asp  
 180 185 190

60

Leu Tyr His Arg Thr Val Ser Arg Leu Gln Tyr His Pro Ser Pro Glu  
 195 200 205

Glu Val Arg Leu Ala Met Arg Asn Ile Pro Ile Glu Ile Thr Gly Glu  
 210 215 220

65

# DE 198 40 771 A 1

Ala Thr Asp Glu Val Glu Val Ser His Arg Asp Ile Pro Gly Val Glu  
225 230 235 240

Thr Asn Gln Leu Arg Gly Gly Ala Ile Leu Val Leu Ala Glu Gly Val  
245 250 255

Leu Gln Lys Ala Lys Lys Leu Val Lys Tyr Ile Asp Lys Met Gly Ile  
260 265 270

Glu Gly Trp Glu Trp Leu Lys Glu Phe Val Glu Ala Lys Glu Lys Gly  
275 280 285

Glu Glu Ile Glu Glu Glu Gly Ser Ala Glu Ser Thr Val Glu Glu Thr  
290 295 300

Lys Val Glu Val Asp Met Gly Phe Tyr Tyr Ser Leu Tyr Gln Lys Phe  
305 310 315 320

Lys Ser Glu Ile Ala Pro Asn Asp Lys Tyr Ala Lys Glu Ile Ile Gly  
325 330 335

Gly Arg Pro Leu Phe Ser Asp Pro Ser Arg Asn Gly Gly Phe Arg Leu  
340 345 350

Arg Tyr Gly Arg Ser Arg Val Ser Gly Phe Ala Thr Trp Gly Ile Asn  
355 360 365

Pro Ala Thr Met Ile Leu Val Asp Glu Phe Leu Ala Ile Gly Thr Gln  
370 375 380

Leu Lys Thr Glu Arg Pro Gly Lys Gly Ala Val Val Thr Pro Val Thr  
385 390 395 400

Thr Ile Glu Gly Pro Ile Val Lys Leu Lys Asp Gly Ser Val Val Lys  
405 410 415

Val Asp Asp Tyr Lys Leu Ala Leu Lys Ile Arg Asp Glu Val Glu Glu  
420 425 430

Ile Leu Tyr Leu Gly Asp Ala Val Ile Ala Phe Gly Asp Phe Val Glu  
435 440 445

Asn Asn Gln Thr Leu Leu Pro Ala Asn Tyr Cys Glu Glu Trp Trp Ile  
450 455 460

Leu Glu Phe Thr Lys Ala Leu Asn Glu Ile Tyr Glu Val Glu Leu Lys  
465 470 475 480

# DE 198 40 771 A 1

Pro Phe Glu Val Asn Ser Ser Glu Asp Leu Glu Glu Ala Ala Asp Tyr  
485 490 495

5 Leu Glu Val Asp Ile Glu Phe Leu Lys Glu Leu Leu Lys Asp Pro Leu  
500 505 510

10 Arg Thr Lys Pro Pro Val Glu Leu Ala Ile His Phe Ser Glu Ile Leu  
515 520 525

15 Gly Ile Pro Leu His Pro Tyr Tyr Thr Leu Tyr Trp Asn Ser Val Lys  
530 535 540

20 Pro Glu Gln Val Glu Lys Leu Trp Arg Val Leu Lys Glu His Ala His  
545 550 555 560

Ile Asp Trp Asp Asn Phe Arg Gly Ile Lys Phe Ala Arg Arg Ile Val  
565 570 575

25 Ile Pro Leu Glu Lys Leu Arg Asp Ser Lys Arg Ala Leu Glu Leu Leu  
580 585 590

30 Gly Leu Pro His Lys Val Glu Gly Lys Asn Val Ile Val Asp Tyr Pro  
595 600 605

35 Trp Ala Ala Ala Leu Leu Thr Pro Leu Gly Asn Leu Glu Trp Glu Phe  
610 615 620

Arg Ala Lys Pro Leu His Thr Thr Ile Asp Ile Ile Asn Glu Asn Asn  
625 630 635 640

40 Glu Ile Lys Leu Arg Asp Arg Gly Ile Ser Trp Ile Gly Ala Arg Met  
645 650 655

45 Gly Arg Pro Glu Lys Ala Lys Glu Arg Lys Met Lys Pro Pro Val Gln  
660 665 670

50 Val Leu Phe Pro Ile Gly Leu Ala Gly Gly Ser Ser Arg Asp Ile Lys  
675 680 685

55 Lys Ala Ala Glu Glu Gly Lys Val Ala Glu Val Glu Ile Ala Leu Phe  
690 695 700

Lys Cys Pro Lys Cys Gly His Val Gly Pro Glu His Ile Cys Pro Asn  
705 710 715 720

60 Cys Gly Thr Arg Lys Glu Leu Ile Trp Val Cys Pro Arg Cys Asn Ala  
725 730 735

65



# DE 198 40 771 A 1

Glu Tyr Pro Glu Ser Gln Ala Ser Gly Tyr Asn Tyr Thr Cys Pro Lys	
740 745 750	
Cys Asn Val Lys Leu Lys Pro Tyr Ala Lys Arg Lys Ile Lys Pro Ser	5
755 760 765	
Glu Leu Leu Lys Arg Ala Met Asp Asn Val Lys Val Tyr Gly Ile Asp	10
770 775 780	
Lys Leu Lys Gly Val Met Gly Met Thr Ser Gly Trp Lys Met Pro Glu	15
785 790 795 800	
Pro Leu Glu Lys Gly Leu Leu Arg Ala Lys Asn Asp Val Tyr Val Phe	20
805 810 815	
Lys Asp Gly Thr Ile Arg Phe Asp Ala Thr Asp Ala Pro Ile Thr His	25
820 825 830	
Phe Arg Pro Arg Glu Ile Gly Val Ser Val Glu Lys Leu Arg Glu Leu	30
835 840 845	
Gly Tyr Thr His Asp Phe Glu Gly Asn Pro Leu Val Ser Glu Asp Gln	35
850 855 860	
Ile Val Glu Leu Lys Pro Gln Asp Ile Ile Leu Ser Lys Glu Ala Gly	40
865 870 875 880	
Lys Tyr Leu Leu Lys Val Ala Lys Phe Val Asp Asp Leu Leu Glu Lys	45
885 890 895	
Phe Tyr Gly Leu Pro Arg Phe Tyr Asn Ala Glu Lys Met Glu Asp Leu	50
900 905 910	
Ile Gly His Leu Val Ile Gly Leu Ala Pro His Thr Ser Ala Gly Ile	55
915 920 925	
Val Gly Arg Ile Ile Gly Phe Val Asp Ala Leu Val Gly Tyr Ala His	60
930 935 940	
Pro Tyr Phe His Ala Ala Lys Arg Arg Asn Cys Phe Pro Gly Asp Thr	65
945 950 955 960	
Arg Ile Leu Val Gln Ile Asn Gly Thr Pro Gln Arg Val Thr Leu Lys	
965 970 975	
Glu Leu Tyr Glu Leu Phe Asp Glu Glu His Tyr Glu Ser Met Val Tyr	
980 985 990	

# DE 198 40 771 A 1

Val Arg Lys Lys Pro Lys Val Asp Ile Lys Val Tyr Ser Phe Asn Pro  
995 1000 1005

5 Glu Glu Gly Lys Val Val Leu Thr Asp Ile Glu Glu Val Ile Lys Ala  
1010 1015 1020

10 Pro Ala Thr Asp His Leu Ile Arg Phe Glu Leu Glu Leu Gly Ser Ser  
1025 1030 1035 1040

15 Phe Glu Thr Thr Val Asp His Pro Val Leu Val Tyr Glu Asn Gly Lys  
1045 1050 1055

Phe Val Glu Lys Arg Ala Phe Glu Val Arg Glu Gly Asn Ile Ile Ile  
1060 1065 1070

20 Ile Ile Asp Glu Ser Thr Leu Glu Pro Leu Lys Val Ala Val Lys Lys  
1075 1080 1085

25 Ile Glu Phe Ile Glu Pro Pro Glu Asp Phe Val Phe Ser Leu Asn Ala  
1090 1095 1100

30 Lys Lys Tyr His Thr Val Ile Ile Asn Glu Asn Ile Val Thr His Gln  
1105 1110 1115 1120

35 Cys Asp Gly Asp Glu Asp Ala Val Met Leu Leu Leu Asp Ala Leu Leu  
1125 1130 1135

Asn Phe Ser Arg Tyr Tyr Leu Pro Glu Lys Arg Gly Gly Lys Met Asp  
1140 1145 1150

40 Ala Pro Leu Val Ile Thr Thr Arg Leu Asp Pro Arg Glu Val Asp Ser  
1155 1160 1165

45 Glu Val His Asn Met Asp Ile Val Arg Tyr Tyr Pro Leu Glu Phe Tyr  
1170 1175 1180

50 Glu Ala Thr Tyr Glu Leu Lys Ser Pro Lys Glu Leu Val Gly Val Ile  
1185 1190 1195 1200

55 Glu Arg Val Glu Asp Arg Leu Gly Lys Pro Glu Met Tyr Tyr Gly Leu  
1205 1210 1215

Lys Phe Thr His Asp Thr Asp Asp Ile Ala Leu Gly Pro Lys Met Ser  
1220 1225 1230

60 Leu Tyr Lys Gln Leu Gly Asp Met Glu Glu Lys Val Lys Arg Gln Leu  
1235 1240 1245

65

# DE 198 40 771 A 1

Asp Val Ala Arg Arg Ile Arg Ala Val Asp Glu His Lys Val Ala Glu  
1250 1255 1260

Thr Ile Leu Asn Ser His Leu Ile Pro Asp Leu Arg Gly Asn Leu Arg  
1265 1270 1275 1280

Ser Phe Thr Arg Gln Glu Phe Arg Cys Val Lys Cys Asn Thr Lys Phe  
1285 1290 1295

Arg Arg Pro Pro Leu Asp Gly Lys Cys Pro Ile Cys Gly Gly Lys Ile  
1300 1305 1310

Val Leu Thr Val Ser Lys Gly Ala Ile Glu Lys Tyr Leu Gly Thr Ala  
1315 1320 1325

Lys Met Leu Val Thr Glu Tyr Lys Val Lys Asn Tyr Thr Arg Gln Arg  
1330 1335 1340

Ile Cys Leu Thr Glu Arg Asp Ile Asp Ser Leu Phe Glu Thr Val Phe  
1345 1350 1355 1360

Pro Glu Thr Gln Leu Thr Leu Leu Val Asn Pro Asn Asp Ile Cys Gln  
1365 1370 1375

Arg Ile Ile Met Glu Arg Thr Gly Gly Ser Lys Lys Ser Gly Leu Leu  
1380 1385 1390

Glu Asn Phe Ala Asn Gly Tyr Asn Lys Gly Lys Lys Glu Glu Met Pro  
1395 1400 1405

Lys Lys Gln Arg Lys Lys Glu Gln Glu Lys Ser Lys Lys Arg Lys Val  
1410 1415 1420

Ile Ser Leu Asp Asp Phe Phe Ser Arg Lys  
1425 1430

## (2) ANGABEN ZU SEQ ID NO: 30:

### (i) SEQUENZKENNZEICHEN:

- (A) LÄNGE: 1092 Aminosäuren
- (B) ART: Aminosäure
- (C) STRANGFORM: Einzelstrang
- (D) TOPOLOGIE: linear

### (ii) ART DES MOLEKÜLS: Protein

### (vi) URSPRÜNGLICHE HERKUNFT:

- (A) ORGANISMUS: Methanobacterium thermoautotrophicum

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 30:

5 Met Met Asp Tyr Phe Asn Glu Leu Glu Arg Glu Thr Glu Arg Leu Tyr  
 1 5 10 15  
 10 Glu Ile Ala Arg Lys Ala Arg Ala Arg Gly Leu Asp Val Ser Thr Thr  
 20 25 30  
 15 Pro Glu Ile Pro Leu Ala Lys Asp Leu Ala Glu Arg Val Glu Gly Leu  
 35 40 45  
 Val Gly Pro Glu Gly Ile Ala Arg Arg Ile Lys Glu Leu Glu Gly Asp  
 50 55 60  
 20 Arg Gly Arg Glu Glu Val Ala Phe Gln Ile Ala Ala Glu Ile Ala Ser  
 65 70 75 80  
 25 Gln Ala Val Pro Asp Asp Asp Pro Glu Glu Arg Glu Lys Leu Ala Asp  
 85 90 95  
 30 Gln Ala Leu Arg Thr Ala Leu Ala Ile Leu Thr Glu Gly Val Val Ala  
 100 105 110  
 Ala Pro Leu Glu Gly Ile Ala Arg Val Arg Ile Lys Glu Asn Phe Asp  
 115 120 125  
 35 Lys Ser Arg Tyr Leu Ala Val Tyr Phe Ala Gly Pro Ile Arg Ser Ala  
 130 135 140  
 40 Gly Gly Thr Ala Ala Ala Leu Ser Val Leu Ile Ala Asp Tyr Ile Arg  
 145 150 155 160  
 45 Leu Ala Val Gly Leu Asp Arg Tyr Lys Pro Val Glu Arg Glu Ile Glu  
 165 170 175  
 Arg Tyr Val Glu Glu Val Glu Leu Tyr Glu Ser Glu Val Thr Asn Leu  
 180 185 190  
 50 Gln Tyr Ser Pro Lys Pro Asp Glu Val Arg Leu Ala Ala Ser Lys Ile  
 195 200 205  
 55 Pro Val Glu Val Thr Gly Glu Pro Thr Asp Lys Val Glu Val Ser His  
 210 215 220  
 60 Arg Asp Leu Glu Arg Val Glu Thr Asn Asn Ile Arg Gly Gly Ala Leu  
 225 230 235 240  
 65

# DE 198 40 771 A 1

Leu Ala Met Val Glu Gly Val Ile Gln Lys Ala Pro Lys Val Leu Lys	
245 250 255	
Tyr Ala Lys Gln Leu Lys Leu Glu Gly Trp Asp Trp Leu Glu Lys Phe	5
260 265 270	
Ser Lys Ala Pro Lys Lys Gly Glu Gly Glu Glu Lys Val Val Val Lys	10
275 280 285	
Ala Asp Ser Lys Tyr Val Glu Asp Ile Ile Gly Gly Arg Pro Val Leu	15
290 295 300	
Ala Tyr Pro Ser Glu Lys Gly Ala Phe Arg Leu Arg Tyr Gly Arg Ala	20
305 310 315 320	
Arg Asn Thr Gly Leu Ala Ala Met Gly Val His Pro Ala Thr Met Glu	25
325 330 335	
Leu Leu Gln Phe Leu Ala Val Gly Thr Gln Met Lys Ile Glu Arg Pro	30
340 345 350	
Gly Lys Gly Asn Cys Val Val Pro Val Asp Thr Ile Asp Gly Pro Val	35
355 360 365	
Val Lys Leu Arg Asn Gly Asp Val Ile Arg Ile Glu Asp Ala Glu Thr	40
370 375 380	
Ala Ser Arg Val Arg Ser Glu Val Glu Glu Ile Leu Phe Leu Gly Asp	45
385 390 395 400	
Met Leu Val Ala Phe Gly Glu Phe Leu Arg Asn Asn His Val Leu Met	50
405 410 415	
Pro Ala Gly Trp Cys Glu Glu Trp Trp Ile Gln Thr Ile Leu Ser Ser	55
420 425 430	
Pro Lys Tyr Pro Gly Asp Asp Pro Leu Asn Leu Ser Tyr Tyr Arg Thr	60
435 440 445	
Arg Trp Asn Glu Leu Glu Val Ser Ala Gly Asp Ala Phe Arg Ile Ser	65
450 455 460	
Glu Glu Tyr Asp Val Pro Leu His Pro Arg Tyr Thr Tyr Phe Tyr His	
465 470 475 480	
Asp Val Thr Val Arg Glu Leu Asn Met Leu Arg Glu Trp Leu Asn Thr	
485 490 495	

# DE 198 40 771 A 1

Ser Gln Leu Glu Asp Glu Leu Val Leu Glu Leu Arg Pro Glu Lys Arg  
500 505 510

5 Ile Leu Glu Ile Leu Gly Val Pro His Arg Val Lys Asp Ser Arg Val  
515 520 525

10 Val Ile Gly His Asp Asp Ala His Ala Leu Ile Lys Thr Leu Arg Lys  
530 535 540

Pro Leu Glu Asp Ser Ser Asp Thr Val Glu Ala Leu Asn Arg Val Ser  
15 545 550 555 560

Pro Val Arg Ile Met Lys Lys Ala Pro Thr Tyr Ile Gly Thr Arg Val  
20 565 570 575

Gly Arg Pro Glu Lys Thr Lys Glu Arg Lys Met Arg Pro Ala Pro His  
25 580 585 590

Val Leu Phe Pro Ile Gly Lys Tyr Gly Gly Ser Arg Arg Asn Ile Pro  
595 600 605

30 Asp Ala Ala Lys Lys Gly Ser Ile Thr Val Glu Ile Gly Arg Ala Thr  
610 615 620

Cys Pro Ser Cys Arg Val Ser Ser Met Gln Ser Ile Cys Pro Ser Cys  
35 625 630 635 640

Gly Ser Arg Thr Val Ile Gly Glu Pro Gly Lys Arg Asn Ile Asn Leu  
40 645 650 655

Ala Ala Leu Leu Lys Arg Ala Ala Glu Asn Val Ser Val Arg Lys Leu  
660 665 670

45 Asp Glu Ile Lys Gly Val Glu Gly Met Ile Ser Ala Glu Lys Phe Pro  
675 680 685

Glu Pro Leu Glu Lys Gly Ile Leu Arg Ala Lys Asn Asp Val Tyr Thr  
50 690 695 700

Phe Lys Asp Ala Thr Ile Arg His Asp Ser Thr Asp Leu Pro Leu Thr  
55 705 710 715 720

His Phe Thr Pro Arg Glu Val Gly Val Ser Val Glu Arg Leu Arg Glu  
725 730 735

60 Leu Gly Tyr Thr Arg Asp Cys Tyr Gly Asp Glu Leu Glu Asp Glu Asp  
740 745 750

65

# DE 198 40 771 A 1

Gln Ile Leu Glu Leu Arg Val Gln Asp Val Val Ile Ser Glu Asp Cys	
755 760 765	
Ala Asp Tyr Leu Val Arg Val Ala Asn Phe Val Asp Asp Leu Leu Glu	5
770 775 780	
Arg Phe Tyr Asp Leu Glu Arg Phe Tyr Asn Val Lys Thr Arg Glu Asp	10
785 790 795 800	
Leu Val Gly His Leu Ile Ala Gly Leu Ala Pro His Thr Ser Ala Ala	15
805 810 815	
Val Leu Gly Arg Ile Ile Gly Phe Thr Gly Ala Ser Ala Cys Tyr Ala	20
820 825 830	
His Pro Tyr Phe His Ser Ala Lys Arg Arg Asn Cys Asp Ser Asp Glu	25
835 840 845	
Asp Ser Val Met Leu Leu Leu Asp Ala Leu Leu Asn Phe Ser Lys Ser	30
850 855 860	
Tyr Leu Pro Ser Ser Arg Gly Gly Ser Met Asp Ala Pro Leu Val Leu	35
865 870 875 880	
Ser Thr Arg Ile Asp Pro Glu Glu Ile Asp Asp Glu Ser His Asn Ile	40
885 890 895	
Asp Thr Met Asp Met Ile Pro Leu Glu Val Tyr Glu Arg Ser Phe Asp	45
900 905 910	
His Pro Arg Pro Ser Glu Val Leu Asp Val Ile Asp Asn Val Glu Lys	50
915 920 925	
Arg Leu Gly Lys Pro Glu Gln Tyr Thr Gly Leu Met Phe Ser His Asn	55
930 935 940	
Thr Ser Arg Ile Asp Glu Gly Pro Lys Val Cys Leu Tyr Lys Leu Leu	60
945 950 955 960	
Pro Thr Met Lys Glu Lys Val Glu Ser Gln Ile Thr Leu Ala Glu Lys	65
965 970 975	
Ile Arg Ala Val Asp Gln Arg Ser Val Val Glu Gly Val Leu Met Ser	
980 985 990	
His Phe Leu Pro Asp Met Met Gly Asn Ile Arg Ala Phe Ser Arg Gln	
995 1000 1005	

# DE 198 40 771 A 1

Lys Val Arg Cys Thr Lys Cys Asn Arg Lys Tyr Arg Arg Ile Pro Leu  
1010 1015 1020

Ser Gly Glu Cys Arg Cys Gly Gly Asn Leu Val Leu Thr Val Ser Lys  
1025 1030 1035 1040

Gly Ser Val Ile Lys Tyr Leu Glu Ile Ser Lys Glu Leu Ala Ser Arg  
1045 1050 1055

Tyr Pro Ile Asp Pro Tyr Leu Met Gln Arg Ile Glu Ile Leu Glu Tyr  
1060 1065 1070

Gly Val Asn Ser Leu Phe Glu Ser Asp Arg Ser Lys Gln Ser Ser Leu  
1075 1080 1085

Asp Val Phe Leu  
1090

(2) ANGABEN ZU SEQ ID NO: 31:

(i) SEQUENZKENNZEICHEN:

(A) LÄNGE: 1263 Aminosäuren

(B) ART: Aminosäure

(C) STRANGFORM: Einzelstrang

(D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: Protein

(vi) URSPRÜNGLICHE HERKUNFT:

(A) ORGANISMUS: Pyrococcus furiosus

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 31:

Met Glu Leu Pro Lys Glu Ile Glu Glu Tyr Phe Glu Met Leu Gln Arg  
1 5 10 15

Glu Ile Asp Lys Ala Tyr Glu Ile Ala Lys Lys Ala Arg Ser Gln Gly  
20 25 30

Lys Asp Pro Ser Thr Asp Val Glu Ile Pro Gln Ala Thr Asp Met Ala  
35 40 45

Gly Arg Val Glu Ser Leu Val Gly Pro Pro Gly Val Ala Gln Arg Ile  
50 55 60



# DE 198 40 771 A 1

Arg Glu Leu Leu Lys Glu Tyr Asp Lys Glu Ile Val Ala Leu Lys Ile  
65 70 75 80

Val Asp Glu Ile Ile Glu Gly Lys Phe Gly Asp Phe Gly Ser Lys Glu  
85 90 95

Lys Tyr Ala Glu Gln Ala Val Arg Thr Ala Leu Ala Ile Leu Thr Glu  
100 105 110

Gly Ile Val Ser Ala Pro Leu Glu Gly Ile Ala Asp Val Lys Ile Lys  
115 120 125

Arg Asn Thr Trp Ala Asp Asn Ser Glu Tyr Leu Ala Leu Tyr Tyr Ala  
130 135 140

Gly Pro Ile Arg Ser Ser Gly Gly Thr Ala Gln Ala Leu Ser Val Leu  
145 150 155 160

Val Gly Asp Tyr Val Arg Arg Lys Leu Gly Leu Asp Arg Phe Lys Pro  
165 170 175

Ser Gly Lys His Ile Glu Arg Met Val Glu Glu Val Asp Leu Tyr His  
180 185 190

Arg Ala Val Ser Arg Leu Gln Tyr His Pro Ser Pro Asp Glu Val Arg  
195 200 205

Leu Ala Met Arg Asn Ile Pro Ile Glu Ile Thr Gly Glu Ala Thr Asp  
210 215 220

Asp Val Glu Val Ser His Arg Asp Val Glu Gly Val Glu Thr Asn Gln  
225 230 235 240

Leu Arg Gly Gly Ala Ile Leu Val Leu Ala Glu Gly Val Leu Gln Lys  
245 250 255

Ala Lys Lys Leu Val Lys Tyr Ile Asp Lys Met Gly Ile Asp Gly Trp  
260 265 270

Glu Trp Leu Lys Glu Phe Val Glu Ala Lys Glu Lys Gly Glu Glu Ile  
275 280 285

Glu Glu Ser Glu Ser Lys Ala Glu Glu Ser Lys Val Glu Thr Arg Val  
290 295 300

Glu Val Glu Lys Gly Phe Tyr Tyr Lys Leu Tyr Glu Lys Phe Arg Ala  
305 310 315 320

# DE 198 40 771 A 1

Glu Ile Ala Pro Ser Glu Lys Tyr Ala Lys Glu Ile Ile Gly Gly Arg  
 325 330 335  
 5 Pro Leu Phe Ala Gly Pro Ser Glu Asn Gly Gly Phe Arg Leu Arg Tyr  
 340 345 350  
 10 Gly Arg Ser Arg Val Ser Gly Phe Ala Thr Trp Ser Ile Asn Pro Ala  
 355 360 365  
 Thr Met Val Leu Val Asp Glu Phe Leu Ala Ile Gly Thr Gln Met Lys  
 15 370 375 380  
 Thr Glu Arg Pro Gly Lys Gly Ala Val Val Thr Pro Ala Thr Thr Ala  
 385 390 395 400  
 20 Glu Gly Pro Ile Val Lys Leu Lys Asp Gly Ser Val Val Arg Val Asp  
 405 410 415  
 25 Asp Tyr Asn Leu Ala Leu Lys Ile Arg Asp Glu Val Glu Glu Ile Leu  
 420 425 430  
 Tyr Leu Gly Asp Ala Ile Ile Ala Phe Gly Asp Phe Val Glu Asn Asn  
 30 435 440 445  
 Gln Thr Leu Leu Pro Ala Asn Tyr Val Glu Glu Trp Trp Ile Gln Glu  
 35 450 455 460  
 Phe Val Lys Ala Val Asn Glu Ala Tyr Glu Val Glu Leu Arg Pro Phe  
 465 470 475 480  
 40 Glu Glu Asn Pro Arg Glu Ser Val Glu Glu Ala Ala Glu Tyr Leu Glu  
 485 490 495  
 45 Val Asp Pro Glu Phe Leu Ala Lys Met Leu Tyr Asp Pro Leu Arg Val  
 500 505 510  
 Lys Pro Pro Val Glu Leu Ala Ile His Phe Ser Glu Ile Leu Glu Ile  
 50 515 520 525  
 Pro Leu His Pro Tyr Tyr Thr Leu Tyr Trp Asn Thr Val Asn Pro Lys  
 55 530 535 540  
 Asp Val Glu Arg Leu Trp Gly Val Leu Lys Asp Lys Ala Thr Ile Glu  
 545 550 555 560  
 60 Trp Gly Thr Phe Arg Gly Ile Lys Phe Ala Lys Lys Ile Glu Ile Ser  
 565 570 575  
 65

# DE 198 40 771 A 1

Leu Asp Asp Leu Gly Ser Leu Lys Arg Thr Leu Glu Leu Leu Gly Leu  
 580 585 590  
 Pro His Thr Val Arg Glu Gly Ile Val Val Val Asp Tyr Pro Trp Ser  
 595 600 605  
 Ala Ala Leu Leu Thr Pro Leu Gly Asn Leu Glu Trp Glu Phe Lys Ala  
 610 615 620  
 Lys Pro Phe Tyr Thr Val Ile Asp Ile Ile Asn Glu Asn Asn Gln Ile  
 625 630 635 640  
 Lys Leu Arg Asp Arg Gly Ile Ser Trp Ile Gly Ala Arg Met Gly Arg  
 645 650 655  
 Pro Glu Lys Ala Lys Glu Arg Lys Met Lys Pro Pro Val Gln Val Leu  
 660 665 670  
 Phe Pro Ile Gly Leu Ala Gly Gly Ser Ser Arg Asp Ile Lys Lys Ala  
 675 680 685  
 Ala Glu Glu Gly Lys Ile Ala Glu Val Glu Ile Ala Phe Phe Lys Cys  
 690 695 700  
 Pro Lys Cys Gly His Val Gly Pro Glu Thr Leu Cys Pro Glu Cys Gly  
 705 710 715 720  
 Ile Arg Lys Glu Leu Ile Trp Thr Cys Pro Lys Cys Gly Ala Glu Tyr  
 725 730 735  
 Thr Asn Ser Gln Ala Glu Gly Tyr Ser Tyr Ser Cys Pro Lys Cys Asn  
 740 745 750  
 Val Lys Leu Lys Pro Phe Thr Lys Arg Lys Ile Lys Pro Ser Glu Leu  
 755 760 765  
 Leu Asn Arg Ala Met Glu Asn Val Lys Val Tyr Gly Val Asp Lys Leu  
 770 775 780  
 Lys Gly Val Met Gly Met Thr Ser Gly Trp Lys Ile Ala Glu Pro Leu  
 785 790 795 800  
 Glu Lys Gly Leu Leu Arg Ala Lys Asn Glu Val Tyr Val Phe Lys Asp  
 805 810 815  
 Gly Thr Ile Arg Phe Asp Ala Thr Asp Ala Pro Ile Thr His Phe Arg  
 820 825 830

5  
 10  
 15  
 20  
 25  
 30  
 35  
 40  
 45  
 50  
 55  
 60  
 65

# DE 198 40 771 A 1

Pro Arg Glu Ile Gly Val Ser Val Glu Lys Leu Arg Glu Leu Gly Tyr  
835 840 845

Thr His Asp Phe Glu Gly Lys Pro Leu Val Ser Glu Asp Gln Ile Val  
850 855 860

Glu Leu Lys Pro Gln Asp Val Ile Leu Ser Lys Glu Ala Gly Lys Tyr  
865 870 875 880

Leu Leu Arg Val Ala Arg Phe Val Asp Asp Leu Leu Glu Lys Phe Tyr  
885 890 895

Gly Leu Pro Arg Phe Tyr Asn Ala Glu Lys Met Glu Asp Leu Ile Gly  
900 905 910

His Leu Val Ile Gly Leu Ala Pro His Thr Ser Ala Gly Ile Val Gly  
915 920 925

Arg Ile Ile Gly Phe Val Asp Ala Leu Val Gly Tyr Ala His Pro Tyr  
930 935 940

Phe His Ala Ala Lys Arg Arg Asn Cys Asp Gly Asp Glu Asp Ser Val  
945 950 955 960

Met Leu Leu Leu Asp Ala Leu Leu Asn Phe Ser Arg Tyr Tyr Leu Pro  
965 970 975

Glu Lys Arg Gly Gly Lys Met Asp Ala Pro Leu Val Ile Thr Thr Arg  
980 985 990

Leu Asp Pro Arg Glu Val Asp Ser Glu Val His Asn Met Asp Val Val  
995 1000 1005

Arg Tyr Tyr Pro Leu Glu Phe Tyr Glu Ala Thr Tyr Glu Leu Lys Ser  
1010 1015 1020

Pro Lys Glu Leu Val Arg Val Ile Glu Gly Val Glu Asp Arg Leu Gly  
1025 1030 1035 1040

Lys Pro Glu Met Tyr Tyr Gly Ile Lys Phe Thr His Asp Thr Asp Asp  
1045 1050 1055

Ile Ala Leu Gly Pro Lys Met Ser Leu Tyr Lys Gln Leu Gly Asp Met  
1060 1065 1070

Glu Glu Lys Val Lys Arg Gln Leu Thr Leu Ala Glu Arg Ile Arg Ala  
1075 1080 1085

# DE 198 40 771 A 1

Val Asp Gln His Tyr Val Ala Glu Thr Ile Leu Asn Ser His Leu Ile  
1090 1095 1100

Pro Asp Leu Arg Gly Asn Leu Arg Ser Phe Thr Arg Gln Glu Phe Arg  
1105 1110 1115 1120

Cys Val Lys Cys Asn Thr Lys Tyr Arg Arg Pro Pro Leu Asp Gly Lys  
1125 1130 1135

Cys Pro Val Cys Gly Gly Lys Ile Val Leu Thr Val Ser Lys Gly Ala  
1140 1145 1150

Ile Glu Lys Tyr Leu Gly Thr Ala Lys Met Leu Val Ala Asn Tyr Asn  
1155 1160 1165

Val Lys Pro Tyr Thr Arg Gln Arg Ile Cys Leu Thr Glu Lys Asp Ile  
1170 1175 1180

Asp Ser Leu Phe Glu Tyr Leu Phe Pro Glu Ala Gln Leu Thr Leu Ile  
1185 1190 1195 1200

Val Asp Pro Asn Asp Ile Cys Met Lys Met Ile Lys Glu Arg Thr Gly  
1205 1210 1215

Glu Thr Val Gln Gly Gly Leu Leu Glu Asn Phe Asn Ser Ser Gly Asn  
1220 1225 1230

Asn Gly Lys Lys Ile Glu Lys Lys Glu Lys Lys Ala Lys Glu Lys Pro  
1235 1240 1245

Lys Lys Lys Lys Val Ile Ser Leu Asp Asp Phe Phe Ser Lys Arg  
1250 1255 1260

(2) ANGABEN ZU SEQ ID NO: 32:

(i) SEQUENZKENNZEICHEN:

(A) LÄNGE: 363 Aminosäuren

(B) ART: Aminosäure

(C) STRANGFORM: Einzelstrang

(D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: Protein

(vi) URSPRÜNGLICHE HERKUNFT:

(A) ORGANISMUS: Homo sapiens

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 32:

5 Met Gln Ala Phe Leu Lys Gly Thr Ser Ile Ser Thr Lys Pro Pro Leu  
 1 5 10 15  
 Thr Lys Asp Arg Gly Val Ala Ala Ser Ala Gly Ser Ser Gly Glu Asn  
 10 20 25 30  
 Lys Lys Ala Lys Pro Val Pro Trp Val Glu Lys Tyr Arg Pro Lys Cys  
 15 35 40 45  
 Val Asp Glu Val Ala Phe Gln Glu Glu Val Val Ala Val Leu Lys Lys  
 50 55 60  
 20 Ser Leu Glu Gly Ala Asp Leu Pro Asn Leu Leu Phe Tyr Gly Pro Pro  
 65 70 75 80  
 Gly Thr Gly Lys Thr Ser Thr Ile Leu Ala Ala Ala Arg Glu Leu Phe  
 25 85 90 95  
 Gly Pro Glu Leu Phe Arg Leu Arg Val Leu Glu Leu Asn Ala Ser Asp  
 30 100 105 110  
 Glu Arg Gly Ile Gln Val Val Arg Glu Lys Val Lys Asn Phe Ala Gln  
 115 120 125  
 35 Leu Thr Val Ser Gly Ser Arg Ser Asp Gly Lys Pro Cys Pro Pro Phe  
 130 135 140  
 40 Lys Ile Val Ile Leu Asp Glu Ala Asp Ser Met Thr Ser Ala Ala Gln  
 145 150 155 160  
 Ala Ala Leu Arg Arg Thr Met Glu Lys Glu Ser Lys Thr Thr Arg Phe  
 45 165 170 175  
 Cys Leu Ile Cys Asn Tyr Val Ser Arg Ile Ile Glu Pro Leu Thr Ser  
 50 180 185 190  
 Arg Cys Ser Lys Phe Arg Phe Lys Pro Leu Ser Asp Lys Ile Gln Gln  
 195 200 205  
 55 Gln Arg Leu Leu Asp Ile Ala Lys Lys Glu Asn Val Lys Ile Ser Asp  
 210 215 220  
 60 Glu Gly Ile Ala Tyr Leu Val Lys Val Ser Glu Gly Asp Leu Arg Lys  
 225 230 235 240  
 65

# DE 198 40 771 A 1

Ala Ile Thr Phe Leu Gln Ser Ala Thr Arg Leu Thr Gly Gly Lys Glu  
245 250 255

Ile Thr Glu Lys Val Ile Thr Asp Ile Ala Gly Val Ile Pro Ala Glu  
260 265 270

Lys Ile Asp Gly Val Phe Ala Ala Cys Gln Ser Gly Ser Phe Asp Lys  
275 280 285

Leu Glu Ala Val Val Lys Asp Leu Ile Asp Glu Gly His Ala Ala Thr  
290 295 300

Gln Leu Val Asn Gln Leu His Asp Val Val Val Glu Asn Asn Leu Ser  
305 310 315 320

Asp Lys Gln Lys Ser Ile Ile Thr Glu Lys Leu Ala Glu Val Asp Lys  
325 330 335

Cys Leu Ala Asp Gly Ala Asp Glu His Leu Gln Leu Ile Ser Leu Cys  
340 345 350

Ala Thr Val Met Gln Gln Leu Ser Gln Asn Cys  
355 360

## (2) ANGABEN ZU SEQ ID NO: 33:

### (i) SEQUENZKENNZEICHEN:

- (A) LÄNGE: 329 Aminosäuren
- (B) ART: Aminosäure
- (C) STRANGFORM: Einzelstrang
- (D) TOPOLOGIE: linear

### (ii) ART DES MOLEKÜLS: Protein

### (vi) URSPRÜNGLICHE HERKUNFT:

- (A) ORGANISMUS: Homo sapiens

### (xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 33:

Asn Leu Val Gln Cys Gly Asp Phe Pro His Leu Leu Val Tyr Gly Pro  
1 5 10 15

Ser Gly Ala Gly Lys Lys Thr Arg Ile Met Cys Ile Leu Arg Glu Leu  
20 25 30

# DE 198 40 771 A 1

Tyr Gly Val Gly Val Glu Lys Leu Arg Ile Glu His Gln Thr Ile Thr  
 35 40 45  
 5 Thr Pro Ser Lys Lys Lys Ile Glu Ile Ser Thr Ile Ala Ser Asn Tyr  
 50 55 60  
 10 His Leu Glu Val Asn Pro Ser Asp Ala Gly Asn Ser Asp Arg Val Val  
 65 70 75 80  
 Ile Gln Glu Met Leu Lys Thr Val Ala Gln Ser Gln Gln Leu Glu Thr  
 15 85 90 95  
 Asn Ser Gln Arg Asp Phe Lys Val Val Leu Leu Thr Glu Val Asp Lys  
 100 105 110  
 20 Leu Thr Lys Asp Ala Gln His Ala Leu Arg Arg Thr Met Glu Lys Tyr  
 115 120 125  
 25 Met Ser Thr Cys Arg Leu Ile Leu Cys Cys Asn Ser Thr Ser Lys Val  
 130 135 140  
 30 Ile Pro Pro Ile Arg Ser Arg Cys Leu Ala Val Arg Val Pro Ala Pro  
 145 150 155 160  
 Ser Ile Glu Asp Ile Cys His Val Leu Ser Thr Val Cys Lys Lys Glu  
 35 165 170 175  
 Gly Leu Asn Leu Pro Ser Gln Leu Ala His Arg Leu Ala Glu Lys Ser  
 180 185 190  
 40 Cys Arg Asn Leu Arg Lys Ala Leu Leu Met Cys Glu Ala Cys Arg Val  
 195 200 205  
 45 Gln Gln Tyr Pro Phe Thr Ala Asp Gln Glu Ile Pro Glu Thr Asp Trp  
 210 215 220  
 Glu Val Tyr Leu Arg Glu Thr Ala Asn Ala Ile Val Ser Gln Gln Thr  
 50 225 230 235 240  
 Pro Gln Arg Leu Leu Glu Val Arg Gly Arg Leu Tyr Glu Leu Leu Thr  
 55 245 250 255  
 His Cys Ile Pro Pro Glu Ile Ile Met Lys Gly Leu Leu Ser Glu Leu  
 260 265 270  
 60 Leu His Asn Cys Asp Gly Gln Leu Lys Gly Glu Val Ala Gln Met Ala  
 275 280 285  
 65



# DE 198 40 771 A 1

Ala Tyr Tyr Glu His Arg Leu Gln Leu Gly Ser Lys Ala Ile Tyr His  
290 295 300

Leu Glu Ala Phe Val Ala Lys Phe Met Ala Leu Tyr Lys Lys Phe Ile  
305 310 315 320

Gln Asp Gly Leu Glu Gly Met Met Phe  
325

## (2) ANGABEN ZU SEQ ID NO: 34:

### (i) SEQUENZKENNZEICHEN:

- (A) LÄNGE: 354 Aminosäuren
- (B) ART: Aminosäure
- (C) STRANGFORM: Einzelstrang
- (D) TOPOLOGIE: linear

### (ii) ART DES MOLEKÜLS: Protein

### (vi) URSPRÜNGLICHE HERKUNFT:

- (A) ORGANISMUS: Homo sapiens

### (xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 34:

Met Glu Val Glu Ala Val Cys Gly Gly Ala Gly Glu Val Glu Ala Gln  
1 5 10 15

Asp Ser Asp Pro Ala Pro Ala Phe Ser Lys Ala Pro Gly Ser Ala Gly  
20 25 30

His Tyr Glu Leu Pro Trp Val Glu Lys Tyr Arg Pro Val Lys Leu Asn  
35 40 45

Glu Ile Val Gly Asn Glu Asp Thr Val Ser Arg Leu Glu Val Phe Ala  
50 55 60

Arg Glu Gly Asn Val Pro Asn Ile Ile Ile Ala Gly Pro Pro Gly Thr  
65 70 75 80

Gly Lys Thr Thr Ser Ile Leu Cys Leu Ala Arg Ala Leu Leu Gly Pro  
85 90 95

Ala Leu Lys Asp Ala Met Leu Glu Leu Asn Ala Ser Asn Asp Arg Gly  
100 105 110

# DE 198 40 771 A 1

Ile Asp Val Val Arg Asn Lys Ile Lys Met Phe Ala Gln Gln Lys Val  
115 120 125

5 Thr Leu Pro Lys Gly Arg His Lys Ile Ile Ile Leu Asp Glu Ala Asp  
130 135 140

10 Ser Met Thr Asp Gly Ala Gln Gln Ala Leu Arg Arg Thr Met Glu Ile  
145 150 155 160

15 Tyr Ser Lys Thr Thr Arg Phe Ala Leu Ala Cys Asn Ala Ser Asp Lys  
165 170 175

Ile Ile Glu Pro Ile Gln Ser Arg Cys Ala Val Leu Arg Tyr Thr Lys  
180 185 190

20 Leu Thr Asp Ala Gln Ile Leu Thr Arg Leu Met Asn Val Ile Glu Lys  
195 200 205

25 Glu Arg Val Pro Tyr Thr Asp Asp Gly Leu Glu Ala Ile Ile Phe Thr  
210 215 220

30 Ala Gln Gly Asp Met Arg Gln Ala Leu Asn Asn Leu Gln Ser Thr Phe  
225 230 235 240

35 Ser Gly Phe Gly Phe Ile Asn Ser Glu Asn Val Phe Lys Val Cys Asp  
245 250 255

Glu Pro His Pro Leu Leu Val Lys Glu Met Ile Gln His Cys Val Asn  
260 265 270

40 Ala Asn Ile Asp Glu Ala Tyr Lys Ile Leu Ala His Leu Trp His Leu  
275 280 285

45 Gly Tyr Ser Pro Glu Asp Ile Ile Gly Asn Ile Phe Arg Val Cys Lys  
290 295 300

50 Thr Phe Gln Met Ala Glu Tyr Leu Lys Leu Glu Phe Ile Lys Glu Ile  
305 310 315 320

Gly Tyr Thr His Met Lys Ile Ala Glu Gly Val Asn Ser Leu Leu Gln  
325 330 335

Met Ala Gly Leu Leu Ala Arg Leu Cys Gln Lys Thr Met Ala Pro Val  
340 345 350

60 Ala Ser

65

## (2) ANGABEN ZU SEQ ID NO: 35:

## (i) SEQUENZKENNZEICHEN:

(A) LÄNGE: 366 Aminosäuren 5

(B) ART: Aminosäure

(C) STRANGFORM: Einzelstrang

(D) TOPOLOGIE: linear 10

## (ii) ART DES MOLEKÜLS: Protein

## (vi) URSPRÜNGLICHE HERKUNFT: 15

(A) ORGANISMUS: Escherichia coli

## (xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 35: 20

Met Lys Phe Thr Val Glu Arg Glu His Leu Leu Lys Pro Leu Gln Gln  
 1 5 10 15 25

Val Ser Gly Pro Leu Gly Gly Arg Pro Thr Leu Pro Ile Leu Gly Asn  
 20 25 30 35

Leu Leu Leu Gln Val Ala Asp Gly Thr Leu Ser Leu Thr Gly Thr Asp  
 35 40 45 50

Leu Glu Met Glu Met Val Ala Arg Val Ala Leu Val Gln Pro His Glu  
 50 55 60 65

Pro Gly Ala Thr Thr Val Pro Ala Arg Lys Phe Phe Asp Ile Cys Arg  
 65 70 75 80 85

Gly Leu Pro Glu Gly Ala Glu Ile Ala Val Gln Leu Glu Gly Glu Arg  
 85 90 95 100

Met Leu Val Arg Ser Gly Arg Ser Arg Phe Ser Leu Ser Thr Leu Pro  
 100 105 110 115

Ala Ala Asp Phe Pro Asn Leu Asp Asp Trp Gln Ser Glu Val Glu Phe  
 115 120 125 130

Thr Leu Pro Gln Ala Thr Met Lys Arg Leu Ile Glu Ala Thr Gln Phe  
 130 135 140 145

Ser Met Ala His Gln Asp Val Arg Tyr Tyr Leu Asn Gly Met Leu Phe  
 145 150 155 160 65

# DE 198 40 771 A 1

Glu Thr Glu Gly Glu Glu Leu Arg Thr Val Ala Thr Asp Gly His Arg  
165 170 175

Leu Ala Val Cys Ser Met Pro Ile Gly Gln Ser Leu Pro Ser His Ser  
180 185 190

Val Ile Val Pro Arg Lys Gly Val Ile Glu Leu Met Arg Met Leu Asp  
195 200 205

Gly Gly Asp Asn Pro Leu Arg Val Gln Ile Gly Ser Asn Asn Ile Arg  
210 215 220

Ala His Val Gly Asp Phe Ile Phe Thr Ser Lys Leu Val Asp Gly Arg  
225 230 235 240

Phe Pro Asp Tyr Arg Arg Val Leu Pro Lys Asn Pro Asp Lys His Leu  
245 250 255

Glu Ala Gly Cys Asp Leu Leu Lys Gln Ala Phe Ala Arg Ala Ala Ile  
260 265 270

Leu Ser Asn Glu Lys Phe Arg Gly Val Arg Leu Tyr Val Ser Glu Asn  
275 280 285

Gln Leu Lys Ile Thr Ala Asn Asn Pro Glu Gln Glu Ala Glu Glu  
290 295 300

Ile Leu Asp Val Thr Tyr Ser Gly Ala Glu Met Glu Ile Gly Phe Asn  
305 310 315 320

Val Ser Tyr Val Leu Asp Val Leu Asn Ala Leu Lys Cys Glu Asn Val  
325 330 335

Arg Met Met Leu Thr Asp Ser Val Ser Ser Val Gln Ile Glu Asp Ala  
340 345 350

Ala Ser Gln Ser Ala Ala Tyr Val Val Met Pro Met Arg Leu  
355 360 365

## (2) ANGABEN ZU SEQ ID NO: 36:

### (i) SEQUENZKENNZEICHEN:

- (A) LÄNGE: 363 Aminosäuren
- (B) ART: Aminosäure
- (C) STRANGFORM: Einzelstrang
- (D) TOPOLOGIE: linear

### (ii) ART DES MOLEKÜLS: Protein

(vi) URSPRÜNGLICHE HERKUNFT:

(A) ORGANISMUS: Aquifex Aeolicus

5

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 36:

Met Arg Val Lys Val Asp Arg Glu Glu Leu Glu Glu Val Leu Lys Lys  
 1 5 10 15

10

Ala Arg Glu Ser Thr Glu Lys Lys Ala Ala Leu Pro Ile Leu Ala Asn  
 20 25 30

15

Phe Leu Leu Ser Ala Lys Glu Glu Asn Leu Ile Val Arg Ala Thr Asp  
 35 40 45

20

Leu Glu Asn Tyr Leu Val Val Ser Val Lys Gly Glu Val Glu Glu Glu  
 50 55 60

25

Gly Glu Val Cys Val His Ser Gln Lys Leu Tyr Asp Ile Val Lys Asn  
 65 70 75 80

Leu Asn Ser Ala Tyr Val Tyr Leu His Thr Glu Gly Glu Lys Leu Val  
 85 90 95

30

Ile Thr Gly Gly Lys Ser Thr Tyr Lys Leu Pro Thr Ala Pro Ala Glu  
 100 105 110

35

Asp Phe Pro Glu Phe Pro Glu Ile Val Glu Gly Gly Glu Thr Leu Ser  
 115 120 125

40

Gly Asn Leu Leu Val Asn Gly Ile Glu Lys Val Glu Tyr Ala Ile Ala  
 130 135 140

45

Lys Glu Glu Ala Asn Ile Ala Leu Gln Gly Met Tyr Leu Arg Gly Tyr  
 145 150 155 160

Glu Asp Arg Ile His Phe Val Gly Ser Asp Gly His Arg Leu Ala Leu  
 165 170 175

50

Tyr Glu Pro Leu Gly Glu Phe Ser Lys Glu Leu Leu Ile Pro Arg Lys  
 180 185 190

55

Ser Leu Lys Val Leu Lys Lys Leu Ile Thr Gly Ile Glu Asp Val Asn  
 195 200 205

60

Ile Glu Lys Ser Glu Asp Glu Ser Phe Ala Tyr Phe Ser Thr Pro Glu  
 210 215 220

65

DE 198 40 771 A 1

Trp Lys Leu Ala Val Arg Leu Leu Glu Gly Glu Phe Pro Asp Tyr Met  
225 230 235 240

Ser Val Ile Pro Glu Glu Phe Ser Ala Glu Val Leu Phe Glu Thr Glu  
245 250 255

Glu Val Leu Lys Val Leu Lys Arg Leu Lys Ala Leu Ser Glu Gly Lys  
260 265 270

Val Phe Pro Val Lys Ile Thr Leu Ser Glu Asn Leu Ala Ile Phe Glu  
275 280 285

Phe Ala Asp Pro Glu Phe Gly Glu Ala Arg Glu Glu Ile Glu Val Glu  
290 295 300

Tyr Thr Gly Glu Pro Phe Glu Ile Gly Phe Asn Gly Lys Tyr Leu Met  
305 310 315 320

Glu Ala Leu Asp Ala Tyr Asp Ser Glu Arg Val Trp Phe Lys Phe Thr  
325 330 335

Thr Pro Asp Thr Ala Thr Leu Leu Glu Ala Glu Asp Tyr Glu Lys Glu  
340 345 350

Pro Tyr Lys Cys Ile Ile Met Pro Met Arg Val  
355 360

(2) ANGABEN ZU SEQ ID NO: 37:

(i) SEQUENZKENNZEICHEN:

- (A) LÄNGE: 1160 Aminosäuren
- (B) ART: Aminosäure
- (C) STRANGFORM: Einzelstrang
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: Protein

(vi) URSPRÜNGLICHE HERKUNFT:

- (A) ORGANISMUS: Escherichia coli

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 37:

Met Ser Glu Pro Arg Phe Val His Leu Arg Val His Ser Asp Tyr Ser  
1 5 10 15

# DE 198 40 771 A 1

Met Ile Asp Gly Leu Ala Lys Thr Ala Pro Leu Val Lys Lys Ala Ala  
20 25 30

Ala Leu Gly Met Pro Ala Leu Ala Ile Thr Asp Phe Thr Asn Leu Cys  
35 40 45

Gly Leu Val Lys Phe Tyr Gly Ala Gly His Gly Ala Gly Ile Lys Pro  
50 55 60

Ile Val Gly Ala Asp Phe Asn Val Gln Cys Asp Leu Leu Gly Asp Glu  
65 70 75 80

Leu Thr His Leu Thr Val Leu Ala Ala Asn Asn Thr Gly Tyr Gln Asn  
85 90 95

Leu Thr Leu Leu Ile Ser Lys Ala Tyr Gln Arg Gly Tyr Gly Ala Ala  
100 105 110

Gly Pro Ile Ile Asp Arg Asp Trp Leu Ile Glu Leu Asn Glu Gly Leu  
115 120 125

Ile Leu Leu Ser Gly Gly Arg Met Gly Asp Val Gly Arg Ser Leu Leu  
130 135 140

Arg Gly Asn Ser Ala Leu Val Asp Glu Cys Val Ala Phe Tyr Glu Glu  
145 150 155 160

His Phe Pro Asp Arg Tyr Phe Leu Glu Leu Ile Arg Thr Gly Arg Pro  
165 170 175

Asp Glu Glu Ser Tyr Leu His Ala Ala Val Glu Leu Ala Glu Ala Arg  
180 185 190

Gly Leu Pro Val Val Ala Thr Asn Asp Val Arg Phe Ile Asp Ser Ser  
195 200 205

Asp Phe Asp Ala His Glu Ile Arg Val Ala Ile His Asp Gly Phe Thr  
210 215 220

Leu Asp Asp Pro Lys Arg Pro Arg Asn Tyr Ser Pro Gln Gln Tyr Met  
225 230 235 240

Arg Ser Glu Glu Glu Met Cys Glu Leu Phe Ala Asp Ile Pro Glu Ala  
245 250 255

Leu Ala Asn Thr Val Glu Ile Ala Lys Arg Cys Asn Val Thr Val Arg  
260 265 270

# DE 198 40 771 A 1

Leu Gly Glu Tyr Phe Leu Pro Gln Phe Pro Thr Gly Asp Met Ser Thr  
 275 280 285

5 Glu Asp Tyr Leu Val Lys Arg Ala Lys Glu Gly Leu Glu Glu Arg Leu  
 290 295 300

10 Ala Phe Leu Phe Pro Asp Glu Glu Glu Arg Leu Lys Arg Arg Pro Glu  
 305 310 315 320

15 Tyr Asp Glu Arg Leu Glu Thr Glu Leu Gln Val Ile Asn Gln Met Gly  
 325 330 335

20 Phe Pro Gly Tyr Phe Leu Ile Val Met Glu Phe Ile Gln Trp Ser Lys  
 340 345 350

25 Asp Asn Gly Val Pro Val Gly Pro Gly Arg Gly Ser Gly Ala Gly Ser  
 355 360 365

30 Leu Val Ala Tyr Ala Leu Lys Ile Thr Asp Leu Asp Pro Leu Glu Phe  
 370 375 380

35 Asp Leu Leu Phe Glu Arg Phe Leu Asn Pro Glu Arg Val Ser Met Pro  
 385 390 395 400

40 Asp Phe Asp Val Asp Phe Cys Met Glu Lys Arg Asp Gln Val Ile Glu  
 405 410 415

45 His Val Ala Asp Met Tyr Gly Arg Asp Ala Val Ser Gln Ile Ile Thr  
 420 425 430

50 Phe Gly Thr Met Ala Ala Lys Ala Val Ile Arg Asp Val Gly Arg Val  
 435 440 445

55 Leu Gly His Pro Tyr Gly Phe Val Asp Arg Ile Ser Lys Leu Ile Pro  
 450 455 460

60 Pro Asp Pro Gly Met Thr Leu Ala Lys Ala Phe Glu Ala Glu Pro Gln  
 465 470 475 480

65 Leu Pro Glu Ile Tyr Glu Ala Asp Glu Glu Val Lys Ala Leu Ile Asp  
 485 490 495

70 Met Ala Arg Lys Leu Glu Gly Val Thr Arg Asn Ala Gly Lys His Ala  
 500 505 510

75 Gly Gly Val Val Ile Ala Pro Thr Lys Ile Thr Asp Phe Ala Pro Leu  
 515 520 525



# DE 198 40 771 A 1

Tyr Cys Asp Glu Glu Gly Lys His Pro Val Thr Gln Phe Asp Lys Ser  
530 535 540

Asp Val Glu Tyr Ala Gly Leu Val Lys Phe Asp Phe Leu Gly Leu Arg  
545 550 555 560

Thr Leu Thr Ile Ile Asn Trp Ala Leu Glu Met Ile Asn Lys Arg Arg  
565 570 575

Ala Lys Asn Gly Glu Pro Pro Leu Asp Ile Ala Ala Ile Pro Leu Asp  
580 585 590

Asp Lys Lys Ser Phe Asp Met Leu Gln Arg Ser Glu Thr Thr Ala Val  
595 600 605

Phe Gln Leu Glu Ser Arg Gly Met Lys Asp Leu Ile Lys Arg Leu Gln  
610 615 620

Pro Asp Cys Phe Glu Asp Met Ile Ala Leu Val Ala Leu Phe Arg Pro  
625 630 635 640

Gly Pro Leu Gln Ser Gly Met Val Asp Asn Phe Ile Asp Arg Lys His  
645 650 655

Gly Arg Glu Glu Ile Ser Tyr Pro Asp Val Gln Trp Gln His Glu Ser  
660 665 670

Leu Lys Pro Val Leu Glu Pro Thr Tyr Gly Ile Ile Leu Tyr Gln Glu  
675 680 685

Gln Val Met Gln Ile Ala Gln Val Leu Ser Gly Tyr Thr Leu Gly Gly  
690 695 700

Ala Asp Met Leu Arg Arg Ala Met Gly Lys Lys Lys Pro Glu Glu Met  
705 710 715 720

Ala Lys Gln Arg Ser Val Phe Ala Glu Gly Ala Glu Lys Asn Gly Ile  
725 730 735

Asn Ala Glu Leu Ala Met Lys Ile Phe Asp Leu Val Glu Lys Phe Ala  
740 745 750

Gly Tyr Gly Phe Asn Lys Ser His Ser Ala Ala Tyr Ala Leu Val Ser  
755 760 765

Tyr Gln Thr Leu Trp Leu Lys Ala His Tyr Pro Ala Glu Phe Met Ala  
770 775 780

# DE 198 40 771 A 1

Ala Val Met Thr Ala Asp Met Asp Asn Thr Glu Lys Val Val Gly Leu  
785 790 795 800

5 Val Asp Glu Cys Trp Arg Met Gly Leu Lys Ile Leu Pro Pro Asp Ile  
805 810 815

10 Asn Ser Gly Leu Tyr His Phe His Val Asn Asp Asp Gly Glu Ile Val  
820 825 830

Tyr Gly Ile Gly Ala Ile Lys Gly Val Gly Glu Gly Pro Ile Glu Ala  
15 835 840 845

Ile Ile Glu Ala Arg Asn Lys Gly Gly Tyr Phe Arg Glu Leu Phe Asp  
20 850 855 860

Leu Cys Ala Arg Thr Asp Thr Lys Lys Leu Asn Arg Arg Val Leu Glu  
865 870 875 880

25 Lys Leu Ile Met Ser Gly Ala Phe Asp Arg Leu Gly Pro His Arg Ala  
885 890 895

30 Ala Leu Met Asn Ser Leu Gly Asp Ala Leu Lys Ala Ala Asp Gln His  
900 905 910

Ala Lys Ala Glu Ala Ile Gly Gln Ala Asp Met Phe Gly Val Leu Ala  
35 915 920 925

Glu Glu Pro Glu Gln Ile Glu Gln Ser Tyr Ala Ser Cys Gln Pro Trp  
930 935 940

40 Pro Glu Gln Val Val Leu Asp Gly Glu Arg Glu Thr Leu Gly Leu Tyr  
945 950 955 960

45 Leu Thr Gly His Pro Ile Asn Gln Tyr Leu Lys Glu Ile Glu Arg Tyr  
965 970 975

Val Gly Gly Val Arg Leu Lys Asp Met His Pro Thr Glu Arg Gly Lys  
50 980 985 990

Val Ile Thr Ala Ala Gly Leu Val Val Ala Ala Arg Val Met Val Thr  
55 995 1000 1005

Lys Arg Gly Asn Arg Ile Gly Ile Cys Thr Leu Asp Asp Arg Ser Gly  
1010 1015 1020

60 Arg Leu Glu Val Met Leu Phe Thr Asp Ala Leu Asp Lys Tyr Gln Gln  
1025 1030 1035 1040

65

# DE 198 40 771 A 1

Leu Leu Glu Lys Asp Arg Ile Leu Ile Val Ser Gly Gln Val Ser Phe  
1045 1050 1055

Asp Asp Phe Ser Gly Gly Leu Lys Met Thr Ala Arg Glu Val Met Asp  
1060 1065 1070

Ile Asp Glu Ala Arg Glu Lys Tyr Ala Arg Gly Leu Ala Ile Ser Leu  
1075 1080 1085

Thr Asp Arg Gln Ile Asp Asp Gln Leu Leu Asn Arg Leu Arg Gln Ser  
1090 1095 1100

Leu Glu Pro His Arg Ser Gly Thr Ile Pro Val His Leu Tyr Tyr Gln  
1105 1110 1115 1120

Arg Ala Asp Ala Arg Ala Arg Leu Arg Phe Gly Ala Thr Trp Arg Val  
1125 1130 1135

Ser Pro Ser Asp Arg Leu Leu Asn Asp Leu Arg Gly Leu Ile Gly Ser  
1140 1145 1150

Glu Gln Val Glu Leu Glu Phe Asp  
1155 1160

## (2) ANGABEN ZU SEQ ID NO: 38:

### (i) SEQUENZKENNZEICHEN:

- (A) LÄNGE: 1161 Aminosäuren
- (B) ART: Aminosäure
- (C) STRANGFORM: Einzelstrang
- (D) TOPOLOGIE: linear

### (ii) ART DES MOLEKÜLS: Protein

### (vi) URSPRÜNGLICHE HERKUNFT:

- (A) ORGANISMUS: Aquifex Aeolicus

### (xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 38:

Met Ser Lys Asp Phe Val His Leu His Leu His Thr Gln Phe Ser Leu  
1 5 10 15

Leu Asp Gly Ala Ile Lys Ile Asp Glu Leu Val Lys Lys Ala Lys Glu  
20 25 30

# DE 198 40 771 A 1

Tyr Gly Tyr Lys Ala Val Gly Met Ser Asp His Gly Asn Leu Phe Gly  
 35 40 45  
 5  
 Ser Tyr Lys Phe Tyr Lys Ala Leu Lys Ala Glu Gly Ile Lys Pro Ile  
 50 55 60  
 10 Ile Gly Met Glu Ala Tyr Phe Thr Thr Gly Ser Arg Phe Asp Arg Lys  
 65 70 75 80  
 15 Thr Lys Thr Ser Glu Asp Asn Ile Thr Asp Lys Tyr Asn His His Leu  
 85 90 95  
 20 Ile Leu Ile Ala Lys Asp Asp Lys Gly Leu Lys Asn Leu Met Lys Leu  
 100 105 110  
 Ser Thr Leu Ala Tyr Lys Glu Gly Phe Tyr Tyr Lys Pro Arg Ile Asp  
 115 120 125  
 25 Tyr Glu Leu Leu Glu Lys Tyr Gly Glu Gly Leu Ile Ala Leu Thr Ala  
 130 135 140  
 30 Cys Leu Lys Gly Val Pro Thr Tyr Tyr Ala Ser Ile Asn Glu Val Lys  
 145 150 155 160  
 35 Lys Ala Glu Glu Trp Val Lys Lys Phe Lys Asp Ile Phe Gly Asp Asp  
 165 170 175  
 40 Leu Tyr Leu Glu Leu Gln Ala Asn Asn Ile Pro Glu Gln Glu Val Ala  
 180 185 190  
 Asn Arg Asn Leu Ile Glu Ile Ala Lys Lys Tyr Asp Val Lys Leu Ile  
 195 200 205  
 45 Ala Thr Gln Asp Ala His Tyr Leu Asn Pro Glu Asp Arg Tyr Ala His  
 210 215 220  
 50 Thr Val Leu Met Ala Leu Gln Met Lys Lys Thr Ile His Glu Leu Ser  
 225 230 235 240  
 55 Ser Gly Asn Phe Lys Cys Ser Asn Glu Asp Leu His Phe Ala Pro Pro  
 245 250 255  
 60 Glu Tyr Met Trp Lys Lys Phe Glu Gly Lys Phe Glu Gly Trp Glu Lys  
 260 265 270  
 Ala Leu Leu Asn Thr Leu Glu Val Met Glu Lys Thr Ala Asp Ser Phe  
 275 280 285  
 65

# DE 198 40 771 A 1

Glu Ile Phe Glu Asn Ser Thr Tyr Leu Leu Pro Lys Tyr Asp Val Pro  
290 295 300

Pro Asp Lys Thr Leu Glu Glu Tyr Leu Arg Glu Leu Ala Tyr Lys Gly  
305 310 315 320

Leu Arg Gln Arg Ile Glu Arg Gly Gln Ala Lys Asp Thr Lys Glu Tyr  
325 330 335

Trp Glu Arg Leu Glu Tyr Glu Leu Glu Val Ile Asn Lys Met Gly Phe  
340 345 350

Ala Gly Tyr Phe Leu Ile Val Gln Asp Phe Ile Asn Trp Ala Lys Lys  
355 360 365

Asn Asp Ile Pro Val Gly Pro Gly Arg Gly Ser Ala Gly Gly Ser Leu  
370 375 380

Val Ala Tyr Ala Ile Gly Ile Thr Asp Val Asp Pro Ile Lys His Gly  
385 390 395 400

Phe Leu Phe Glu Arg Phe Leu Asn Pro Glu Arg Val Ser Met Pro Asp  
405 410 415

Ile Asp Val Asp Phe Cys Gln Asp Asn Arg Glu Lys Val Ile Glu Tyr  
420 425 430

Val Arg Asn Lys Tyr Gly His Asp Asn Val Ala Gln Ile Ile Thr Tyr  
435 440 445

Asn Val Met Lys Ala Lys Gln Thr Leu Arg Asp Val Ala Arg Ala Met  
450 455 460

Gly Leu Pro Tyr Ser Thr Ala Asp Lys Leu Ala Lys Leu Ile Pro Gln  
465 470 475 480

Gly Asp Val Gln Gly Thr Trp Leu Ser Leu Glu Glu Met Tyr Lys Thr  
485 490 495

Pro Val Glu Glu Leu Leu Gln Lys Tyr Gly Glu His Arg Thr Asp Ile  
500 505 510

Glu Asp Asn Val Lys Lys Phe Arg Gln Ile Cys Glu Glu Ser Pro Glu  
515 520 525

Ile Lys Gln Leu Val Glu Thr Ala Leu Lys Leu Glu Gly Leu Thr Arg  
530 535 540

# DE 198 40 771 A 1

His Thr Ser Leu His Ala Ala Gly Val Val Ile Ala Pro Lys Pro Leu  
 545 550 555 560

5 Ser Glu Leu Val Pro Leu Tyr Tyr Asp Lys Glu Gly Glu Val Ala Thr  
 565 570 575

10 Gln Tyr Asp Met Val Gln Leu Glu Glu Leu Gly Leu Leu Lys Met Asp  
 580 585 590

15 Phe Leu Gly Leu Lys Thr Leu Thr Glu Leu Lys Leu Met Lys Glu Leu  
 595 600 605

20 Ile Lys Glu Arg His Gly Val Asp Ile Asn Phe Leu Glu Leu Pro Leu  
 610 615 620

25 Asp Asp Pro Lys Val Tyr Lys Leu Leu Gln Glu Gly Lys Thr Thr Gly  
 625 630 635 640

30 Val Phe Gln Leu Glu Ser Arg Gly Met Lys Glu Leu Leu Lys Lys Leu  
 645 650 655

35 Lys Pro Asp Ser Phe Asp Asp Ile Val Ala Val Leu Ala Leu Tyr Arg  
 660 665 670

40 Pro Gly Pro Leu Lys Ser Gly Leu Val Asp Thr Tyr Ile Lys Arg Lys  
 675 680 685

45 His Gly Lys Glu Pro Val Glu Tyr Pro Phe Pro Glu Leu Glu Pro Val  
 690 695 700

50 Leu Lys Glu Thr Tyr Gly Val Ile Val Tyr Gln Glu Gln Val Met Lys  
 705 710 715 720

55 Met Ser Gln Ile Leu Ser Gly Phe Thr Pro Gly Glu Ala Asp Thr Leu  
 725 730 735

60 Arg Lys Ala Ile Gly Lys Lys Lys Ala Asp Leu Met Ala Gln Met Lys  
 740 745 750

65 Asp Lys Phe Ile Gln Gly Ala Val Glu Arg Gly Tyr Pro Glu Glu Lys  
 755 760 765

70 Ile Arg Lys Leu Trp Glu Asp Ile Glu Lys Phe Ala Ser Tyr Ser Phe  
 770 775 780

75 Asn Lys Ser His Ser Val Ala Tyr Gly Tyr Ile Ser Tyr Trp Thr Ala  
 785 790 795 800

# DE 198 40 771 A 1

Tyr Val Lys Ala His Tyr Pro Ala Glu Phe Phe Ala Val Lys Leu Thr	
805 810 815	
Thr Glu Lys Asn Asp Asn Lys Phe Leu Asn Leu Ile Lys Asp Ala Lys	5
820 825 830	
Leu Phe Gly Phe Glu Ile Leu Pro Pro Asp Ile Asn Lys Ser Asp Val	10
835 840 845	
Gly Phe Thr Ile Glu Gly Glu Asn Arg Ile Arg Phe Gly Leu Ala Arg	15
850 855 860	
Ile Lys Gly Val Gly Glu Glu Thr Ala Lys Ile Ile Val Glu Ala Arg	20
865 870 875 880	
Lys Lys Tyr Lys Gln Phe Lys Gly Leu Ala Asp Phe Ile Asn Lys Thr	25
885 890 895	
Lys Asn Arg Lys Ile Asn Lys Lys Val Val Glu Ala Leu Val Lys Ala	30
900 905 910	
Gly Ala Phe Asp Phe Thr Lys Lys Lys Arg Lys Glu Leu Leu Ala Lys	35
915 920 925	
Val Ala Asn Ser Glu Lys Ala Leu Met Ala Thr Gln Asn Ser Leu Phe	40
930 935 940	
Gly Ala Pro Lys Glu Glu Val Glu Glu Leu Asp Pro Leu Lys Leu Glu	45
945 950 955 960	
Lys Glu Val Leu Gly Phe Tyr Ile Ser Gly His Pro Leu Asp Asn Tyr	50
965 970 975	
Glu Lys Leu Leu Lys Asn Arg Tyr Thr Pro Ile Glu Asp Leu Glu Glu	55
980 985 990	
Trp Asp Lys Glu Ser Glu Ala Val Leu Thr Gly Val Ile Thr Glu Leu	60
995 1000 1005	
Lys Val Lys Lys Thr Lys Asn Gly Asp Tyr Met Ala Val Phe Asn Leu	65
1010 1015 1020	
Val Asp Lys Thr Gly Leu Ile Glu Cys Val Val Phe Pro Gly Val Tyr	
1025 1030 1035 1040	
Glu Glu Ala Lys Glu Leu Ile Glu Glu Asp Arg Val Val Val Val Lys	
1045 1050 1055	

Gly Phe Leu Asp Glu Asp Leu Glu Thr Glu Asn Val Lys Phe Val Val  
 1060 1065 1070

5 Lys Glu Val Phe Ser Pro Glu Glu Phe Ala Lys Glu Met Arg Asn Thr  
 1075 1080 1085

10 Leu Tyr Ile Phe Leu Lys Arg Glu Gln Ala Leu Asn Gly Val Ala Glu  
 1090 1095 1100

15 Lys Leu Lys Gly Ile Ile Glu Asn Asn Arg Thr Glu Asp Gly Tyr Asn  
 1105 1110 1115 1120

20 Leu Val Leu Thr Val Asp Leu Gly Asp Tyr Phe Val Asp Leu Ala Leu  
 1125 1130 1135

Pro Gln Asp Met Lys Leu Lys Ala Asp Arg Lys Val Val Glu Glu Ile  
 1140 1145 1150

25 Glu Lys Leu Gly Val Lys Val Ile Ile  
 1155 1160

30 Patentansprüche

1. Thermostabiler in vitro Komplex zur Template-abhängigen Elongation von Nukleinsäuren, umfassend ein thermostabiles Gleitklammerprotein, welches mit einem thermostabilen Polymeraseaktivität-aufweisenden Elongationsprotein verbunden ist.

35 2. Thermostabiler Komplex nach Anspruch 1, dadurch gekennzeichnet, dass das Gleitklammerprotein und das Elongationsprotein über ein Kopplungsprotein verbunden sind.

3. Thermostabiler Komplex nach Anspruch 1 oder 2, dadurch gekennzeichnet, dass das Gleitklammerprotein und das Elongationsprotein aus Archaeobakterien stammen.

40 4. Thermostabiler Komplex nach Anspruch 1, 2 oder 3, dadurch gekennzeichnet, dass das Gleitklammerprotein eine ringförmige Struktur aufweist, welche die Template-Nukleinsäurestränge ganz oder teilweise umschließt.

5. Thermostabiler Komplex nach einem der Ansprüche 1 bis 4, dadurch gekennzeichnet, dass das Gleitklammerprotein eine oder beide der folgenden Aminosäurekonsensussequenzen aufweist:

SEQ ID NO: 39

45 [G A V L I M P F W]-D-X-X-X-[G A V L I M P F W]-X-X-[G A V L I M P F W]-X-[G A V L I M P F W]-X-[G A V L I M P F W]-X-X-X-X-F-X-X-Y-X-X-D und/oder

SEQ ID NO: 40

[G A V L I M P F W]-X(3)-L-A-P-[K R H D E]-[G A V L I M P F W]-E.

6. Thermostabiler Komplex nach einem der Ansprüche 1 bis 5, dadurch gekennzeichnet, dass das Gleitklammerprotein zu der menschlichen (Eukaryonten) PCNA Aminosäuresequenz (SEQ ID NO: 11) auf einer Länge von mindestens 100 Aminosäuren bei einem Sequenzalignment eine mindestens 20%ige Sequenzidentität aufweist und/oder das Gleitklammerprotein zu der bakterielle  $\beta$ -clamp Sequenz aus E.coli (Eubakteria) (SEQ ID NO: 35) auf einer Länge von mindestens 100 Aminosäuren bei einem Sequenzalignment eine mindestens 20%ige Sequenzidentität aufweist und/oder das Gleitklammerprotein zu der Aminosäuresequenz des PCNA Homologen aus Archaeoglobus Fulgidus (Archaea) (SEQ ID NO: 12) auf einer Länge von mindestens 100 Aminosäuren bei einem Sequenzalignment eine mindestens 20%ige Sequenzidentität aufweist.

7. Thermostabiler Komplex nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, dass das Gleitklammerprotein mit einem aus einem Alignment aus Abb. 12 generierten Hidden Markow Modell einen Score von mindestens 20 ergibt und/oder wobei das Gleitklammerprotein mit einem aus einem Alignment aus Abb. 13 generierten Hidden Markow Modell einen Score von mindestens 25 ergibt.

8. Thermostabiler Komplex nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, dass das Gleitklammerprotein ausgewählt ist aus AF0335 aus Archaeoglobus Fulgidus, MJ0247 aus Methanococcus jannaschii, PHLA008 aus Pyrococcus horikoshii, MTH1312 aus Methanobacterium thermoautotrophicus sowie AE000761\_7 aus Aquifex aeolicus.

9. Thermostabiler Komplex nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, dass das Elongationsprotein eine 5'-3'-Polymeraseaktivität oder eine Reverse-Transkriptaseaktivität aufweist.

10. Thermostabiler Komplex nach Anspruch 9, dadurch gekennzeichnet, dass das Elongationsprotein mindestens eine der folgenden Konsensussequenzen beinhaltet und an nicht mehr als vier Positionen von dieser Sequenz abweicht:



SEQ ID NO: 44

D-[G A V L I M P F W]-[G A V L I M P F W]-X-X-Y-N-X-X-X-F-D-X-P-Y-[G A V K U N O F W]-X-X-R-A

SEQ ID NO: 45

A-[G A V L I M P F W]-R-T-A-[F A V L I M P F W]-A-[G A V L I M P F W]-[G A V L I M P F W]-T-E-G-[G A V L I M P F W]-V-X-A-P-[G A V L I M P F W]-E-G-I-A-X-V-[K R H D E]-I

SEQ ID NO: 46

[G A V L I M P F W]-P-V-G-[G A V L I M P F W]-G-R-G-S-X-[G A V L I M P F W]-G-S-[G A V K U N O F W]-V-A-X-A-[G A V L I M P F W]-X-I-T-D-[G A V K U N O F W]-D-P-[G A V L I M P F W]-X-X-X-[G A V L I M P F W]-L-F-E-R-F-L-N-P-E-R-[G A V L I M P F W]-S-M-P-D.

11. Thermostabiler Komplex nach Anspruch 9, dadurch gekennzeichnet, dass das Elongationsprotein eine zu der menschlichen (Eukaryonten) Aminosäuresequenz (SEQ ID NO: 22) auf einer Länge von mindestens 200 Aminosäuren bei einem Sequenzalignment eine mindestens 20%ige Sequenzidentität aufweist und/oder zur archaebakteriellen Aminosäuresequenz (SEQ ID NO: 27) auf einer Länge von mindestens 400 Aminosäuren bei einem Sequenzalignment eine mindestens 25%ige Sequenzidentität aufweist und/oder zur eubakteriellen Aminosäuresequenz (SEQ ID NO: 37) auf einer Länge von mindestens 300 Aminosäuren bei einem Sequenzalignment eine mindestens 25%ige Sequenzidentität aufweist.

12. Thermostabiler Komplex nach Anspruch 9, dadurch gekennzeichnet, dass das Elongationsprotein mit einem aus einem Alignment aus Abb. 17 generierten Hidden Markow Modell einen Score von mindestens 20 ergibt und/oder wobei das Elongationsprotein mit einem aus einem Alignment aus Abb. 18 generierten Hidden Markow Modell einen Score von mindestens 35 ergibt und/oder wobei das Elongationsprotein mit einem aus einem Alignment aus Abb. 19 generierten Hidden Markow Modell einen Score von mindestens 20 ergibt.

13. Thermostabiler Komplex nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, dass das Elongationsprotein ausgewählt ist aus AF0497 oder AF1722 aus *Archaeoglobus fulgidus*, MJ0885 oder MJ1630 aus *Methanococcus jannaschii*, PHBT047 oder PHBN021 aus *Pyrococcus horikoshii*, MTH1208 oder MTH1536 aus *Methanobacterium thermoautotrophicus* sowie PFUORF3 aus *Pyrococcus furiosus*.

14. Thermostabiler Komplex nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, dass das Kopplungsprotein die folgende Konsensussequenz beinhaltet und an nicht mehr als vier Positionen von dieser Sequenz abweicht:

(SEQ ID NO: 43)

[FL]-[G A V L I M P F W]-X-X-[G A V L I M P F W]-X-G-X(13)-[G A V L I M P F W]-X-[YR]-[G A V L I M P F W]-X-X-[G A V L I M P F W]-A-G-[DN]-[G A V L I M P F W]-[G A V L I M P F W]-[DS].

15. Thermostabiler Komplex nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, dass das Kopplungsprotein zu der menschlichen (Eukaryonten) Aminosäuresequenz (SEQ ID NO: 16) auf einer Länge von mindestens 150 Aminosäuren bei einem Sequenzalignment eine mindestens 18%ige Sequenzidentität aufweist.

16. Thermostabiler Komplex nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, dass das Kopplungsprotein mit einem aus einem Alignment aus Abb. 16 generierten Hidden Markow Modell einen Score von mindestens 10 ergibt.

17. Thermostabiler Komplex nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, dass das Kopplungsprotein ausgewählt ist aus AF1790 aus *Archaeoglobus fulgidus*, MJ0702 aus *Methanococcus jannaschii*, PHBN023 aus *Pyrococcus horikoshii*, MTH1405 aus *Methanobacterium thermoautotrophicum* sowie PFUORF2 aus *Pyrococcus furiosus*.

18. Thermostabiler Komplex nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, dass der Komplex mit einem Protein assoziiert ist, welches als Gleitklammerlader fungiert.

19. Thermostabiler Komplex nach einem der vorhergehenden Ansprüche, dadurch gekennzeichnet, dass der Komplex mit ATP oder einem anderen Kofaktor assoziiert vorliegt.

20. Thermostabiler akzessorischer in vitro-Komplex, dadurch gekennzeichnet, dass er ein Gleitklammerprotein und ein Kopplungsprotein enthält, wie sie jeweils in einem der vorhergehenden Ansprüche definiert sind.

21. Rekombinante DNA-Sequenz, dadurch gekennzeichnet, dass sie für einen thermostabilen in vitro-Komplex gemäß einem der Ansprüche 1 bis 19 oder für einen thermostabilen akzessorischen Komplex nach Anspruch 20 codiert.

22. Vektor, dadurch gekennzeichnet, dass er eine für ein Gleitklammerprotein und ein Kopplungsprotein oder/und ein Elongationsprotein codierende, rekombinante DNA-Sequenz enthält.

23. Vektor nach Anspruch 22, dadurch gekennzeichnet, dass er zusätzlich geeignete Restriktionsschnittstellen zur Insertion weiterer DNA-Sequenzen in einer solchen Anordnung enthält, dass sich ein Fusionsprotein aus Gleitklammerprotein und dem Expressionsprodukt der weiteren DNA-Sequenzen ergibt.

24. Vektor nach Anspruch 22 oder 23, dadurch gekennzeichnet, dass er geeignete, die Expression der DNA-Sequenz(en) steuernde Promotor- oder/und Operatorbereiche enthält.

25. Vektor nach Anspruch 24, dadurch gekennzeichnet, dass er mehrere Promotor- oder/und Operatorbereiche zur getrennten Expression mehrerer DNA-Sequenzen enthält.

26. Vektor nach einem der Ansprüche 22 bis 25, dadurch gekennzeichnet, dass er reprimier- und induzierbare Promotor- oder/und Operatorbereiche enthält.

27. Vektor nach einem der Ansprüche 22 bis 26, dadurch gekennzeichnet, dass er eine DNA-Sequenz nach Anspruch 21 enthält.

28. Wirtszelle, dadurch gekennzeichnet, dass sie mit einem oder mehreren Vektoren nach einem der Ansprüche 22 bis 27 transformiert ist.

29. Verfahren zur Herstellung eines thermostabilen in vitro-Komplexes gemäß einem der Ansprüche 1 bis 19 oder eines thermostabilen, akzessorischen in vitro-Komplexes gemäß Anspruch 20, dadurch gekennzeichnet, dass man eine entsprechende rekombinante DNA-Sequenz gemäß Anspruch 21 oder einen oder mehrere entsprechende Vek-

toren gemäß einem der Ansprüche 22 bis 27 in eine Wirtszelle einbringt, die Expression der Proteine bewirkt und diese aus dem Kulturmedium oder nach Zellaufschluss isoliert und gegebenenfalls noch mit weiteren Komplexbestandteilen koppelt.

30. Verwendung eines thermostabilen, akzessorischen in vitro-Komplexes nach Anspruch 20 zur Herstellung eines thermostabilen in vitro-Komplexes nach einem der Ansprüche 1 bis 19, dadurch gekennzeichnet, dass man den akzessorischen Komplex mit einem Polymeraseaktivität aufweisenden Elongationsprotein koppelt.

31. Verfahren zur Template-abhängigen Elongation von Nukleinsäuren, wobei die Nukleinsäure nötigenfalls denaturiert, mit mindestens einem Primer unter Hybridisierungsbedingungen versehen wird, wobei der Primer komplementär zu einem flankierenden Bereich einer gewünschten Nukleinsäuresequenz des Template-Strangs ist und mit Hilfe einer Polymerase in Anwesenheit von Nukleotiden eine Primer-Elongation erfolgt, dadurch gekennzeichnet, dass als Polymerase ein thermostabiler in vitro-Komplex gemäß einem der Ansprüche 1 bis 19 verwendet wird.

32. Verfahren nach Anspruch 31, dadurch gekennzeichnet, dass man zur Amplifikation von DNA-Sequenzen zwei die gewünschte Nukleinsäuresequenz flankierende Primer und Deoxynukleotide oder/und Derivate davon oder/und Ribonukleotide oder/und Derivate davon verwendet.

33. Verfahren nach Anspruch 32, dadurch gekennzeichnet, dass man eine Polymerase-Ketten-Reaktion durchführt.

34. Verfahren nach Anspruch 31, dadurch gekennzeichnet, dass man zur Reversen-Transkription von RNA in DNA einen thermostabilen in vitro-Komplex gemäß einem der Ansprüche 1 bis 19 einsetzt, dessen Elongationsprotein Reverse-Transkriptase-Aktivität aufweist.

35. Verfahren nach einem der Ansprüche 31 bis 34, dadurch gekennzeichnet, dass man zur Sequenzierung von Nukleinsäuren ausgehend von einem Primer, der zu einem der zu sequenzierenden Nukleinsäure benachbarten Bereich komplementär ist, eine Template-abhängige Elongation bzw. Reverse-Transkription unter Verwendung von Deoxynukleotiden und Dideoxynukleotiden oder deren jeweiligen Derivaten gemäß der Methode von Sanger durchführt.

36. Verfahren nach einem der Ansprüche 31 bis 35, dadurch gekennzeichnet, dass man bei der Elongation der Nukleinsäuren Markierungen einfügt.

37. Verfahren nach Anspruch 36, dadurch gekennzeichnet, dass man markierte Primer oder/und markierte Deoxynukleotide oder/und deren Derivate oder/und markierte Dideoxynukleotide oder/und deren Derivate oder/und markierte Ribonukleotide oder/und deren Derivate einsetzt.

38. Verfahren zur Markierung von Nukleinsäuren durch Erzeugung einzelner Brüche in Phosphodiesterbindungen der Nukleinsäurekette und Ersatz eines Nukleotids an den Bruchstellen durch ein markiertes Nukleotid mit Hilfe einer Polymerase, dadurch gekennzeichnet, dass als Polymerase ein thermostabiler in vitro-Komplex nach einem der Ansprüche 1 bis 19 eingesetzt wird.

39. Reagenzien-Kit zur Elongation oder/und Amplifikation oder/und Reversen-Transkription oder/und Sequenzierung oder/und Markierung von Nukleinsäuren, enthaltend in einem oder mehreren getrennten Behältern

a) einen thermostabilen in vitro-Komplex gemäß einem der Ansprüche 1 bis 19 oder

b) einen thermostabilen akzessorischen in vitro-Komplex gemäß Anspruch 20 und gegebenenfalls separat davon ein Polymeraseaktivität aufweisendes Elongationsprotein,

sowie gegebenenfalls Primer, Puffersubstanzen, Nukleotide, ATP, einen oder mehrere andere Kofaktoren oder/und Pyrophosphat.

40. Kit nach Anspruch 39, dadurch gekennzeichnet, dass er zur Amplifikation von Nukleinsäuren neben den Substanzen a) oder b), welche 5'-3'-Polymeraseaktivität aufweisen, Deoxynukleotide oder/und Derivate davon enthält.

41. Kit nach Anspruch 39, dadurch gekennzeichnet, dass er zur Reversen Transkription Substanzen a) oder b) enthält, welche Reverse Transkriptase-Aktivität aufweisen, sowie Deoxynukleotide oder/und Derivate davon.

42. Kit nach einem der Ansprüche 39 bis 41, dadurch gekennzeichnet, dass er zur Sequenzierung neben Deoxynukleotiden oder Ribonukleotiden oder/und deren Derivaten, Dideoxynukleotide oder/und deren Derivate enthält.

43. Kit nach einem der Ansprüche 39 bis 42, dadurch gekennzeichnet, dass er Primer oder/und Deoxynukleotide oder/und Dideoxynukleotide in markierter Form enthält.

Hierzu 15 Seite(n) Zeichnungen

- Leerseite -

Abbildung 1:

Komponenten des Elongations- holoenzyms	Eukaryota Homo sapiens	Archaeon Archaeoglobus Fulgidus (%Identität   Alignmentlänge) <sup>1</sup>	Archaeon Methanococcus Jannaschii	Archaeon Pyrococcus Horikoshii (%Identität) <sup>2</sup>	Archaeon Methanobact. Thermoaerol.	Archaeon Pyrococcus Furiosus (%Identität) <sup>2</sup>
Gleitkamm- lad α I	AC11_HUMAN [SEQ ID NO:1] AC12_HUMAN [SEQ ID NO:32] AC13_HUMAN [SEQ ID NO:33] AC14_HUMAN [SEQ ID NO:34]	AF2060 (39%   307) [SEQ ID NO:2] AF2060 (45%) [SEQ ID NO:2] AF2060 (27%) [SEQ ID NO:2] AF2060 (42%) [SEQ ID NO:2]	MJ1422 (32%   173) [SEQ ID NO:3]	PHBN012 (39%   269) [SEQ ID NO:4]	MTH0241 (35%   317) [SEQ ID NO:5]	-----
Gleitkamm- lad α II	AC15_HUMAN [SEQ ID NO:6]	AF1195 (29%   276) [SEQ ID NO:7]	MJ0884 (25%   571) [SEQ ID NO:8]	PHBN013 (23%   429) [SEQ ID NO:9]	MTH0240 (26%   308) [SEQ ID NO:10]	-----
Gleitkamm α	PCNA_HUMAN [SEQ ID NO:11]	AF0335 (24%   257) [SEQ ID NO:12]	MJ0247 (25%   256) [SEQ ID NO:13]	PHLA008 (25%   245) [SEQ ID NO:14]	MTH1312 (28%   260) [SEQ ID NO:15]	-----
koppelnde Untereinheit	DPD2_HUMAN [SEQ ID NO:16]	AF1790 (23%   211) [SEQ ID NO:17]	MJ0702 (19%   340) [SEQ ID NO:18]	PHBN023 (21%   401) <sup>3</sup> [SEQ ID NO:19]	MTH1405 (21%   265) [SEQ ID NO:20]	PFJOREF2 (43%   378) <sup>2</sup> [SEQ ID NO:21]
Elongationsprot. I (Polymetase)	DPOD_HUMAN [SEQ ID NO:22]	AF0497 (25%   742) [SEQ ID NO:23]	MJ0885 (28%   277) [SEQ ID NO:24]	PHBT047 (29%   439) [SEQ ID NO:25]	MTH1208 (26%   650) [SEQ ID NO:26]	-----
Elongationsprot. II (Pol. Aktivität)	-----	AF1722 [SEQ ID NO:27]	MJ1630 (50%   1160) <sup>2</sup> [SEQ ID NO:28]	PHBN021 (47%   974) <sup>2</sup> [SEQ ID NO:29]	MTH1536 (49%   1139) <sup>2</sup> [SEQ ID NO:30]	PFJOREF3 (50%   1198) <sup>2</sup> [SEQ ID NO:31]

Abbildung 2:

Komponenten des Elongations- holoenzymes	Eubakteria Escherichia coli	Aquifex Aeolicus (%Identität   Alignmentlänge) <sup>1</sup>
Gleitklammer	DP3B_ECOLI [SEQ ID NO:35]	AASEQ93 (28%   370) [SEQ ID NO:36]
Elongationsprotein	DP3A_ECOLI [SEQ ID NO:37]	AASEQ50 (37%   1154) [SEQ ID NO:38]

Abbildung 3:

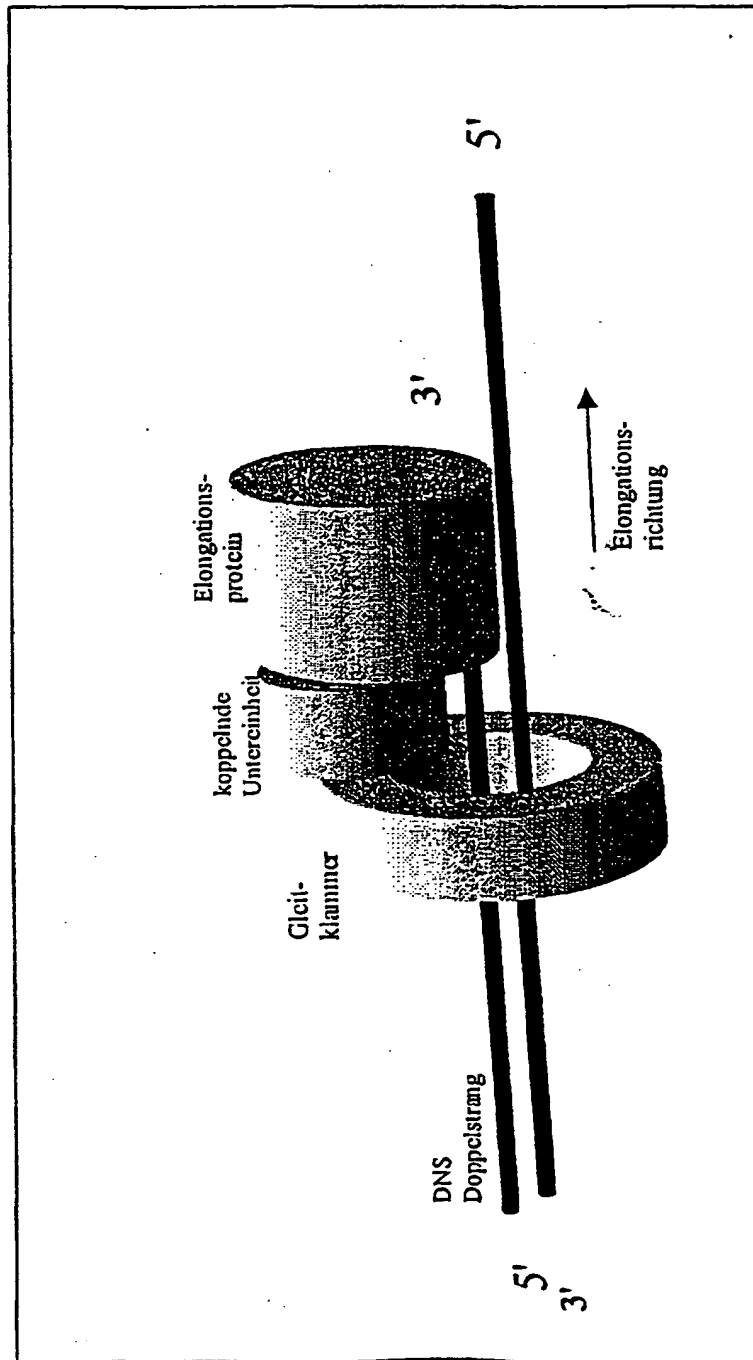


Abbildung 4:

## Gleitklammer Region 1:

PCNA\_HUMAN : MDSSHVSLVQLTLRSEGFDTYRCD  
PCNA\_METJA : MDP SHVALVSLEIPRLAFEEYEAD  
MTH1312 : LDRSHITYVHLELKAELFDEYVCD  
PHLA008 : MDPSRVVLIDLNLPSIFSKYEVD  
AF0335 : VDPANVAMVIVDIPKDSFEVYNID

Konsensus : \$DXXX\$XX\$X\$X\$XXXXFXXYXXD

Pattern : [GAVLIMPFW]-D-X-X-X-[GAVLIMPFW]-X-X-[GAVLIMPFW]-X-  
[GAVLIMPFW]-X-[GAVLIMPFW]-X-X-X-X-F-X-X-Y-X-X-D

## Gleitklammer Region 2:

PCNA\_HUMAN : LKYYLAPKIE  
PCNA\_METJA : LTFL LAPRIE  
MTH1312 : LSFL LAPRIE  
PHLA008 : LTFL LAPRVE  
AF0335 : VEYILAPRIE

Konsensus : \$XXXLAP&\$E

Pattern : [GAVLIMPFW]-X(3)-L-A-P-[KRHDE]-[GAVLIMPFW]-E

Abbildung 5:

Koppelnde Untereinheit:

PHBN023 : FLEWLNGYVESKEEEEIVSRIRYLIIAGDVVD  
 PfuORF2 : FLEWLNGNVETKEEEEIVSRVKYLIIAGDVVD  
 DPD2\_HUMAN: LVDVVTGQLGDEGEQCSAAHVSRVILAGNLLS  
 MTH1405 : FVKWINGDFGSEEQRSLADV KYLVVAGDIVD  
 AF1790 : FVRWLKGEVGGKKSONLAEKV KYI VIAGDIVD  
 MJ0702 : FIRFLNGDVDNELEEKVVSRLKYIC IAGDLVD

Konsensus : F\$XX\$XGXXXXXXXXXXXXX\$XY\$X\$AGD\$D  
 L R N S

Pattern : [FL]-[GAVLIMPFW]-X-X-[GAVLIMPFW]-X-G-X(13)-[GAVLIMPFW]-X-  
 [YR]-[GAVLIMPFW]-X-[GAVLIMPFW]-A-G-[DN]-[GAVLIMPFW]-[GAVLIMPFW]-[DS]

Abbildung 6:

Gleitklammerlader:

AC11\_HUMAN : CNYLSKIIPALQSRCTRFRFGPL  
 AF2060 : CNYVSRIIEPIQSRCAVFRFKPV  
 MTH0241 : CNYSSKIIDPIQSRCAIFRFLPL  
 PHBN012 : CNYSSKIIEPIQSRCAIFRFRPL  
 MJ1422 : CNYPSKIIPPIQSRCAVFRFSPL

Konsensus : CNYXS&IIX\$\$QSRCXXFRFXP\$

Pattern : C-N-Y-X-S-[KRHDE]-I-I-X-[GAVLIMPFW]-[GAVLIMPFW]-Q-S-R-C-  
 X-X-F-R-F-X-P-[GAVLIMPFW]



Abbildung 7:

## Gleitklammerlader 2:

AC15\_HUMAN : KAALLSGPPPGVGKTTTASLV  
MJ0884 : KPILLVGPPPGCGKTTLAYAL  
AF1195 : KPLLLAGPPPGVGKTSLLALAL  
MTH0240 : KPLLLVGPPPGTGKTTLAHII  
PHBN013 : KALLLAGPPPGSGKTTTVYAL

Konsensus : KXXLLXGPPGXGKT\$X\$XX\$

Pattern : K-X-X-L-L-X-G-P-P-G-X-G-K-T-[STNQYC]-X-[GAVLIMPFW]-X-X-[GAVLIMPFW]

Abbildung 8:

## Elongationprotein 1

DPOD\_HUMAN : DVITGYNIQNFDLPYLISRA  
DPOL\_ARCFU : DIIVGYNQDAFDWPYLKRKA  
MTH1208 : DILVGYNSDNFDFFPYITRRA  
PHBT047 : DVIITYNGDNFDFFPYLLKRA  
MJ0885 : DVIITYNGDNFDFFPYLKARA

Konsensus : D\$XXYNXXXFDXPY\$XXRA

Pattern : D-[GAVLIMPFW]-[GAVLIMPFW]-X-X-Y-N-X-X-X-F-D-X-P-Y-[GAVLIMPFW]-X-X-R-A

Abbildung 9:

## Elongationsprotein 2

AF1722 : AVRTAVAIMTEGVVAAPIEGIARVRI  
 MJ1630 : AVRTALAVLTEGIVAAPLEGIADV KI  
 PfuORF3 : AVRTALAILTEGIVSAPLEGIADV KI  
 MTH1536 : ALRTALAILTEGVVAAPLEGIARVRI  
 PHBN021 : AVRTALAILTEGVVSAPIEGIASVKI

Konsensus: A\$RTA\$A\$T\$EG\$V\$XAP\$EGIA\$XV&I

Pattern : A-[GAVLIMPFW]-R-T-A-[GAVLIMPFW]-A-[GAVLIMPFW]-[GAVLIMPFW]-  
 T-E-G-[GAVLIMPFW]-V-X-A-P-[GAVLIMPFW]-E-G-I-A-X-V-[KRHDE]-I

Abbildung 10:

## Elongationsprotein (Eubakteria)

DP3A\_ECOLI : VPVGPGRGSGAGSLVAYALKITDLDPLEFDLLFERFLNPERVSMPD  
 DP3A\_SALTY : VPVGPGRGSGAGSLVAYALKITDLDPLEFDLLFERFLNPERVSMPD  
 BB0579 : IPVGAGRGSGAGSIVAYALRITDIDPLKYNLLFERFLNPERISMPD  
 DP3A\_HELPY : IPVGPGRGSAAGSLVAFALKITDIDPLKYDLLFERFLNPERISMPD  
 AA50 : IPVGPGRGSAGGSLVAYAIGITDVPDIKHGFLFERFLNPERVSMPD

Konsensus : \$PVG\$GRGSX\$GS\$VAXA\$XITD\$DP\$XXX\$LFERFLNPER\$SMPD

Pattern : [GAVLIMPFW]-P-V-G-[GAVLIMPFW]-G-R-G-S-X-[GAVLIMPFW]-G-S-  
 [GAVLIMPFW]-V-A-X-A-[GAVLIMPFW]-X-I-T-D-[GAVLIMPFW]-D-P-[GAVLIMPFW]-X-  
 X-X-[GAVLIMPFW]-L-F-E-R-F-L-N-P-E-R-[GAVLIMPFW]-S-M-P-D

Abbildung 11:

## Gleitklammer (Eubakteria) Region 1:

AAPOL3B : KLVITGGKSTYKLPTAPAEDFP  
DP3B\_ECOLI : RMLVRSGRSRFSLSLTLPAADFP  
S.TYPHIM. : RMLVRSGRSRFSLSLTLPAADFP  
DP3B\_PROMI : RLLVRSGRSRFSLSLTLPASDFP  
DP3B\_PSEPU : KLLVKAGRSRFTLSTLPANDFP  
DP3B\_STRCO : RATVVCSSRFTLHTLPVEEYP

Konsensus : &XX\$XXGX\$XXXLXT\$P\$X&XP

Pattern : [KRHDE]-X-X-[GAVLIMPFW]-X-X-G-X-S-X-X-X-L-X-T-  
[GAVLIMPFW]-P-[GAVLIMPFW]-X-[KRHDE]-X-P

## Gleitklammer (Eubakteria) Region 2:

AAPOL3B : IIMPMRV  
DP3B\_ECOLI : VVMPMRL  
S.TYPHIM. : VVMPMRL  
DP3B\_PROMI : VVMPMRL  
DP3B\_PSEPU : VVMPMRL  
DP3B\_STRCO : LIMPVRL

Konsensus : \$\$MP\$R\$

Pattern : [GAVLIMPFW]-[GAVLIMPFW]-M-P-[GAVLIMPFW]-R-[GAVLIMPFW]

**Abbildung 12:**

```

      *      20      *      40      *
PCNA_HUMAN : LINEACWDISSSGVNLQSMDSHVSLVQLTLRSEGFDTYRCDRNLAMGVNLTSMKIL : 58
PCNA_METJA : LLDEICFEVDEEGIKASAMDP SHVALVSLEIPRLAFEEYEADS-HDIGIDLEAFKKVM : 57
MTH1312 : IVDEVQIQLSAEGLRLDALDRSHITYVHLELKAELFDEYVCDEPERINVDTEELMKVL : 58
PHLA008 : LIDEAAFKVTEEGISMRAMPDSRVVLIDLNLPSISF SKYEVDGEETIGVNMMDHLKKVL : 58
AF0335 : LVSEARIHFLEKGLHSRAVDPANVAMVIVDIPKDSFEVYNIDEETIGVMDMRIFDIS : 58

      60      *      80      *      100      *
PCNA_HUMAN : KCAGNEDIITLRAEDNADTLALVFEAPNQEKVSDYEMKLMDL DVEQLG--IPEQEYSC : 114
PCNA_METJA : NRAKAKDRLLLELDEEKNKLN VIFENTG---KRKFSLALLDISASSVK--VPEIEYPN : 110
MTH1312 : KRAKANDRVILSTDEG-N-LIIQFEGEA---VRTFKIRLIDIEYETPS--PPEIEYEN : 109
PHLA008 : KRGKAKDTLILRK GEE-NFLEISLQGT A---TRTFRLPLIDVEEIEVE--LPDLPYTA : 110
AF0335 : KSISTKDLVELIVEDE-STLKVKFGSVE-----YKVALIDPSAIRKEPRIPELELPA : 109

      120      *      140      *      160      *
PCNA_HUMAN : VVKMPSGEFARICRDL SHIGDAVVISCAKDG VKFSASGELGNGNIKLSQTSNVDKEEE : 172
PCNA_METJA : VIMIKGDAFKEALKDADLFSDYVILKADEK FVIHAKGDLNENEAIFEKDSSA----- : 163
MTH1312 : EFEVPPQLLKDSIADIDIFSDKITFRVDEDRFIASAEGEFGDAQIEY----- : 156
PHLA008 : KVVVLGEVLKEAVKDASLVSDSIKFMAKENEFIMRAEGETQVEV KLTLEDEG----- : 163
AF0335 : KIVMDAGEFKK AIAAADKISDQVIFRS DKEGFRIEAKGDVDSIVFHMT-ET----- : 159

      180      *      200      *      220      *
PCNA_HUMAN : AVTIEMNEPVQLTFALRYLNFFTKATPLSSTVTL SMSADVPLVVEYKIADM-GHLKYY : 229
PCNA_METJA : IISLEVKEEAKSAFNLDYLM DMVKGVS SSGDIKIYLCNDMPLKLEYSIAG--VNLTFL : 219
MTH1312 : LHGERIDKPARSIYSLDKIKEMLKADKFSETAIINLGDDMPLKLT LKMASKEGELSFL : 214
PHLA008 : LLDIEVQEETKSAYGVSYLADMVKGIGKADEV TMRFGNEMPMQMEYYIRDE-GRLTFL : 220
AF0335 : ELIEFNGGEARS MFSDY LKEFCKVAGSGDLLTIHLGTNYPVRLV FELVGGRAKVEYI : 217

PCNA_HUMAN : LAPKIED : 236
PCNA_METJA : LAPRIEG : 226
MTH1312 : LAPRIEA : 221
PHLA008 : LAPRVEE : 227
AF0335 : LAPRIES : 224

```

Abbildung 13:

```

      *           20           *           40           *
AAPOL3B : MRVKVDREELVKKARESTEKKAALPILANFLLSAKEENLIVRATDLENYLVSVK : 58
DP3B_ECOLI : MKFTVEREHLKPLQOVSGPLGGRPTLPILGNLLQVADGTLSTGTDLMEMVARVA : 58
S.TYPHIM. : MKFTVEREHLKPLQOVSGPLGGRPTLPILGNLLQVADGTLSTGTDLMEMVARVT : 58
DP3B_PROMI : MKFIIEREQLKPLQOVSGPLGGRPTLPILGNLLKVNTLSTGTDLMEMMARVS : 58
DP3B_PSEPU : MHFTIQREALLKPLQLVAGVVERRQTLFVLSNVLLVVQCGQLSLTGTDLLEVELVGRVQ : 58
DP3B_STRCO : MKIRVERDVLAEAVAWAARSLPARPPAPVLAGLLKAEQGQLSLSSFDYEV SARVSVE : 58

      60           *           80           *           100           *
AAPOL3B : GEVEEE-GEVCVHSQKLYDIVKNL-NSAYVYLHTEGEKLVITGGKSTYKLPATAEDF : 114
DP3B_ECOLI : LVQPHPEGATTVPARKFFDICRGLPEGAIEIVQLEGERMLVRSGRSRFSLSTLPAADF : 116
S.TYPHIM. : LSQPHPEGATTVPARKFFDICRGLPEGAIEIVQLEGERMLVRSGRSRFSLSTLPAADF : 116
DP3B_PROMI : LSQSHEIGATTVPARKFFDIWRGLPEGAIEISVELDGRLLVRSGRSRFSLSTLPASDF : 116
DP3B_PSEPU : LEEPAEPGEITVPARKLMDICKSLPNDALIDIKVDEQKLLVKAGRSRFTLSTLPANDF : 116
DP3B_STRCO : AEIEEE-GTVLVSGRLLADISRAL-PNRPVEISTDGV RATVVCSSRFTLHTLPVEEY : 114

      120           *           140           *           160           *
AAPOL3B : PEFPEIVE-GGETLSGNLLVNGIEKVEYAIKKEANIALQGMYL RGYEDRIHFVGS DG : 171
DP3B_ECOLI : PNLDDWQSEVEFTLPQATMKRLIEATQFSMAHQDVRYYLNGMLFETEGEELRTVATDG : 174
S.TYPHIM. : PNLDDWQSEVEFTLPQATMKRLIESTQFSMAHQDVRYYLNGMLFETEGSELRTVATDG : 174
DP3B_PROMI : PNLDDWQSEVEFTLPQATMKRLIESTQFSMAHQDVRYYLNGMLFETENTELRTVATDG : 174
DP3B_PSEPU : PTVEEGPGSLTCNLEQSKLRLRIERTSFMAQQDVRYYLNGMLLEVSRNTLRAVSTDG : 174
DP3B_STRCO : PALPQMPE-ATGTVPGEVFASAVQQAIAAGRDDTL PVL TGVRIEIEGDSVTLASTDR : 171

      180           *           200           *           220           *
AAPOL3B : HRLALYEPLGEFSKEL-----LIPRKS LKVLKKLITGIEDVNI EKSE---DESFAYFS : 221
DP3B_ECOLI : HRLAVCSMPIGQSLPS--HSVIVPRKGVIELMRMLDG-GDNPLRVQI---GSNNIRAH : 226
S.TYPHIM. : HRLAVCSMPLEASLPS--HSVIVPRKGVIELMRMLDG-GENPLRVQI---GSNNIRAH : 226
DP3B_PROMI : HRLAVCAMDIGQSLPG--HSVIVPRKGVIELMRMLDGSGESLLQLQI---GSNNIRAH : 227
DP3B_PSEPU : HRLALC SMSAPIEQEDR-HQVIVPRKGILELARLLTD-PEGMVSVIL---GQHHIRAT : 227
DP3B_STRCO : YRFAVREFLWKPENPDISAVALVPAKTLQDTAKALTSGDQVILALSGSGAGEGLIGFE : 229

      240           *           260           *           280           *
AAPOL3B : TPEWKLA VRLLEGEFPDYM SVIPEEFSAEVLFETEVLKVLKRLKALSEGKVPVKIT : 279
DP3B_ECOLI : VGDFIFTSKLVDGRFPDYRRVLPKNPDKHLEAGCDLLKQAFARAAILSNEKFRGVRLY : 284
S.TYPHIM. : VGDFIFTSKLVDGRFPDYRRVLPKNPDKHLEAGCDILKQAFARAAILSNEKFRGVRLY : 284
DP3B_PROMI : VGDFIFTSKLVDGRFPDYRRVLPKNPTKTVIAGCDILKQAFSRAA ILSNEKFRGVRIN : 285
DP3B_PSEPU : TGEFTFTSKLVDGKFPDYERVLPKGGDKLVVGDRQALREAFSRTAIL SNEKYRGIRLQ : 285
DP3B_STRCO : GAGRRITTRLLEGDL PKYKTLFPTEFNSVAVIETAPFVEAVKRVALVA-ERNTPVRLS : 286

      300           *           320           *           340
AAPOL3B : LSENLAIFEADPEFGAEEREEIEVEYTGEPFEIGFNGKYLMEALDAYDSERVWFKFTT : 337
DP3B_ECOLI : VSENQLKITANNPEQEEAEI LDVTSYGAEMEIGFNVSYVLDVNLKNCENVRMMLT- : 341
S.TYPHIM. : VSENQLKITANNPEQEEAEI LDVSYGGTEMEIGFNVSYVLDVNLKNCETVRIMLT- : 341
DP3B_PROMI : LTNGQLKITANNPEQEEAEI DVVQYQGEEMEIGFNVSYLLDVNLTKCEEVKLLLT- : 342
DP3B_PSEPU : LAAGQLKIQANNPEQEEAEIEI SVDYEGSSLEIGFNVSYLLDV LGVMTTEQVRLILS- : 342
DP3B_STRCO : FEQGVLLILEAGSSDDAQAVERVDAQLEGDDISIAFNPTFLLDGLSAIDSPVAQLSFTT : 344

      *           360           *           380
AAPOL3B : ---PDTATLLEAEDYEK-EPYKCIIMP MRV-- : 363
DP3B_ECOLI : ----DSVSSVQIEDAAS-QSAA YVVM PMRL-- : 366
S.TYPHIM. : ----DSVSSVQIEDAAS-QSAA YVVM PMRL-- : 366
DP3B_PROMI : ----DAVSSVQVENVAS-AAAA YVVM PMRL-- : 367
DP3B_PSEPU : ----DSNSSALLQEAGN-DDSS YVVM PMRL-- : 367
DP3B_STRCO : STKPALLSGRPAVDAAEADAYKYLIMPVRLSG : 376

```

## Abbildung 14:

```

      *           20           *           40           *           60
AC11_HUMAN : ICNYLSKIIPALQSRCTRFRFGPLTPELMVPRLEHVVEEEKVDISEDGMKALVTLSSGDM : 60
AF2060      : SCNYVSRIIEPIQSRCAVFRFKVPVPKEAMKKRLLEICEKEGVKITEDGLEALIYISGGDF : 60
MTH0241     : SCNYSSKIIDPIQSRCAIFRFLPLKGHQIIKRLEYIAEKENLEYEAHALETIVYFAEGDL : 60
PHBN012     : SCNYSSKIIEPIQSRCAIFRFRPLRDEDIAKRLRYIAENEGLELTEEGLQAILYIAEGDM : 60
MJ1422      : NCNYPSKIIPPIQSRCAVFRFSPLKKEDIAKKLKEIAEKEGLNLTESGLEAIIYVSEGDM : 60

```

```

      *           80           *           100          *           120
AC11_HUMAN : RRALNILQSTNMAFGKVTEETVYTTCTGHPLKSDIANILDWMLNQDFTTAYRNITELKTLK : 120
AF2060      : RKAINALQGAAAIGEVDADTIYQITATARPEEMTELIQTALKGNFMEARELLDRLMVEY : 120
MTH0241     : RKAINLLQSAASLGEKITESSIYDVVSRARPKDVRKMIKTILDGKFMEARDMLREIMVLQ : 120
PHBN012     : RRAINILQAAAALDKKITDENVMFVASRARPEIDREMMLLALKGNFLKAREKLREILLKQ : 120
MJ1422      : RKAINVLQTAAALSDVIDDEIVYKVSSRARPEEVKQMELALDGKFMEARDLLYKLMVEW : 120

```

```

      *
AC11_HUMAN : GLALHDILTEIHLFVHRVD : 139
AF2060      : GMSGEDIVAQLFREIISMP : 139
MTH0241     : GISGEDMVTQIYQELSRLA : 139
PHBN012     : GLSGEDVLIQMHKEVFNLP : 139
MJ1422      : GMSGEDILNQMFREINSLD : 139

```

## Abbildung 15:

				*	20		*	40		*
AC15_HUMAN	:	LLWVDKYKPTSLKTIIGQQGDQSCANKLLRWLRNWQKSSSEDKKHAAKFGKFSKGDDG	:							58
MJ0884	:	LSWVEKYRPKSLKDVAG---HEKVKEKLKTWIESYLKGET-----	:							37
AF1195	:	MLWVEKYRPKTLEEVVA---DKSIIITRVIKWKAKSWKRG-----	:							35
MTH0240	:	MSWTEKYRPGSFDEVVG---NQKVIAEIKEWIKAWKAGKP-----	:							37
PHBN013	:	VPWIEKYRPRKLSEIVN---QEQALEKVRWIESWLHGNPPK-----	:							39
		60		*	80		*	100		*
AC15_HUMAN	:	SSFKAALLSGPPGVGKTTTASLVCQELGYSYVELNASDTRSKSSLKAIVAESLNNTSI	:							116
MJ0884	:	--PKPILLVGPPGCGKTTLAYALANDYGFEVIELNASDKRNSSAIKKVVGHAATSSSI	:							93
AF1195	:	--SKPLLLAGPPGVGKTSALALANTMGWEAVELNASDQRSWRVIERIVGEGAFNETI	:							91
MTH0240	:	--QKPLLLVGPPGTGKTTLAHIIGKEFS-DTLELNASDRSQDALMRSAGEASATRS	:							92
PHBN013	:	--KKALLLAGPPGSGKTTTVYALAHEYNFVIELNASDERTYNKIARYV-QAAYTMDI	:							94
		120		*	140		*	160		*
AC15_HUMAN	:	KGFYSNGAASSVSTKHALIMDEVDMAGNEDRGGIQELIGLIK-HTKIPIICMCNDRN	:							173
MJ0884	:	FG-----KKF-LIVLDEVDSISGKEDAGGVSELIKVIK-KAKNPIILTANDAY	:							139
AF1195	:	SDEGEF-LSSRIGKLLIILDEVDSIHKKEDVGGEEALIRLIKRPKPAQPLILIANDPY	:							148
MTH0240	:	FN-----HDLKLIILDEVDSIHGNEEDRGGVQAINRIIK-ESRHPMVLANDPY	:							139
PHBN013	:	MG-----KRRKIIFLDEADNIE--P--SGAPEIAKLID-KARNPIIMAANYHW	:							137
		180		*	200		*	220		*
AC15_HUMAN	:	HPKIRSLVHYCFDLRFQRPVEQIKGAMMSIAFKEGLKIPPPAMNEIILGANQDIRQV	:							231
MJ0884	:	APSIRSLLPYVEVIQNLNPVHTNSVYKVLKKAIEKEGLDVEDDKTLKMIAQHSAGDLRSA	:							197
AF1195	:	KLSP-ELRNLCEMINFKRLTKQQVARVLERIALKEGIKVDKSVLLKIAENAGGDLRAA	:							205
MTH0240	:	SKRLQSIKPRCRVLNLRKVHTSSIAAALRRICRAEGIECPDDVLRRELAKRSRGDLRSA	:							197
PHBN013	:	EVPK-EIRDRAELVEYKRLNQRDVISALVRILKREGITVPKEILTEIAKRSSGDLRAA	:							194
		240		*	260		*	280		*
AC15_HUMAN	:	LHNLSMWCARSKALTYDQAKADSHRAKDKIMGPFVDVARKVFAAGEETAHMSLVDSKSD	:							289
MJ0884	:	INDLEALALSGDLSYEAQKLP----DRKREANIFDALRVILKTTHYGIATTALMN--	:							249
AF1195	:	INDFQALAEKKEELKPEDVFLT----KRTQEKDIFRVQMIFKTKNPAVYNEAML--	:							256
MTH0240	:	INDLEAMAEGERIGEELLKMG----EKDATSNLFDVAVRAVLKSRDVSQVREAMR--	:							248
PHBN013	:	INDLQTTIVAGG--YEDAKYVLA----YRDVEKTVFQSLGMVFSDDNAKRAKLALMN--	:							244
		300		*	320		*	340		
AC15_HUMAN	:	LFFHDYSIAPLFVQENYIHKVPVAAGGDMKKHMLLSRAADSIDGDLVDSQIRSKQN	:							347
MJ0884	:	-----VDETPDVVIEWIAENVPKEYEKPEEVARAFEYLSKADRYLGRVMRRQN--YSP	:							300
AF1195	:	-----LDESPEDVIHWVDENLPLEYSG-VELVNAYEALSADIFLGRVRRRQF--YRL	:							306
MTH0240	:	-----VDDDPVLVLEFIAENVPREYEKPNIEISRAYDMLSRADIFFGRAVRTRN--YTY	:							299
PHBN013	:	-----LDMSPDEFLWVDENIPHYMLKPEEMARAYEASRADIYLGRAQRTGN--YSL	:							295
		*	360		*	380		*	400	
AC15_HUMAN	:	WSLLPAQAIYASVLPGLMRGYMTQFPTFPWSWLGKHSSTGKHDRIVQDLALHMSLRTY	:							405
MJ0884	:	WKYATTLMTAGVALSKDEKYRKWTPYSY-PKIFRLLTKTKAEREILNKILKIGEKTH	:							357
AF1195	:	WKYASYLMTVGVOQMKEEPKKGFTYRR-PAVWQMLFQLRQKREMTRKILEKIGKYSH	:							363
MTH0240	:	WRYASELMGPGVALAKDKTYRKVRYTG-SSSFRILGKTRKQSLRDSVAAKMAGKMH	:							356
PHBN013	:	WKYAIIDMMTAGVAVAGTK-KKGFAPKFPY-PNTLKMLAESKEERSIRDSIIKKIMKEMH	:							351
		*	420							
AC15_HUMAN	:	SSKRTVNMDYLSLLR	:							420
MJ0884	:	TSSKRAR-FDLQMLK	:							371
AF1195	:	LSMRKARTEMFPVIK	:							378
MTH0240	:	ISPKVAI-SMFPYME	:							370
PHBN013	:	MSKLEAL-ETMKILR	:							365

## Abbildung 16:

```

      *           20           *           40           *
PHBN023 : IFEVEDQTDVRVKVFLPKDSED-YREALKVLPAVVAFAKGVYSKRG-IFFANRFYLPDV : 56
PfuORF2 : IFEIEDLTGKVKVFLPKDSED-YREAFKVLPAVVAFAKGVYSKRG-ILYANKFYLPDV : 56
DPD2_HUMAN : -LVLEDELQRIKLKGTIDVSK-----LVTGTVLAVFGSVRDDGKFLVEDYCFADLA : 50
MTH1405 : IIELEDDTGEISVVVHNENHKLFEKSEKIVRDEVVGVHGTKKGR--FVVASEIFHPGV : 56
AF1790 : YIRLEDTTGTITCVATGKNAE---VARELLGDEVIGVTGLLKGS--SLYANRIVFPDV : 53
MJ0702 : IVRIEDTEDEATLILPKEKIEAGKIPDDILLDEVIGAIGTVSKSGSSIYVDEIIRPAL : 58

      60           *           80           *           100           *
PHBN023 : PLYRK-QKPPLEEKVYAVLTSDIHVGSK--EFCEKAFIKFLEWLNGYVESKEEEEIVS : 111
PfuORF2 : PLYRR-QKPPLEEKVYAILISDIHVGSK--EFCENAFIKFLEWLNGNVETKEEEEIVS : 111
DPD2_HUMAN : PQKP--APP-LDTRFVLLVSGGLGGGGESLLGT-QLLVDVVTGQLGDEGEQCSAA : 104
MTH1405 : PRIQ--EK---EMDFSVAFTSDVHIGSQ--TFLEDAFMKFVKWINGDFGSEEQRSLAA : 107
AF1790 : PINGNEK---KRDFYIVFLSDTHFGSK--EFLEKEWEMFVRWLKGEVGGKKSQNLAE : 106
MJ0702 : PPKEP-KR--IDEEYMAFLSDIHVGSK--EFLHKEFEKPIRFLNGVDVNELEEKVVS : 111

      120           *           140           *           160           *
PHBN023 : RIRYL_IAGDVVD-GIGIYP-GQYSDLIIPDIFDQYEALANLISNVP----KHITIFI : 163
PfuORF2 : RVKYLI_IAGDVVD-GVGVPY-GQYADLTIPDIFDQYEALANLISNVP----KHITMFI : 163
DPD2_HUMAN : HVS RVILAGNLLSHSTQSRDSINKAKYLTQKQAASVEAVKMLDEILLQLSASVPVDV : 162
MTH1405 : DVKYLVVAGDIVD-GIGIYP-GQEKELLIRDIHEQYEEAARLFGDIR----SDIKIVM : 159
AF1790 : KVKYIIV_IAGDIVD-GIGVYP-GQEDDLAISDIYGQYEFAAASHLDEIP----KEIKIIV : 158
MJ0702 : RLKYICI_IAGDLVD-GVGVPY-GQEDDLYEVDLIEQYREIAMYLDQIP----EHISIII : 163

      180           *           200           *           220           *
PHBN023 : GPGNHDAARPAIQPEFYEEYAKPLYKLKNTVIIISNPAVIRLHGRDFLIAHGRGIEDV : 221
PfuORF2 : APGNHDAARQAIPQPEFYKEYAKPIYKLKNAVIIISNPAVIRLHGRDFLIAHGRGIEDV : 221
DPD2_HUMAN : MPGEFDPTNYTLPOQPLHPCMFPLATAYSTLQLVTNPYQATIDGVRFLGTSGQNVSDI : 220
MTH1405 : IPGNHDSRIAEPQPAIPPEYAKSLYSIRNIEFLSNPSLSLDGVRTLIYHGRSFDDM : 217
AF1790 : SPGNHDAVRQAEPQPAFEGEI-RSLFP-KNVEHVGNPAYVDIEGVKVLIIYHGRSIDI : 214
MJ0702 : SPGNHDAVRPAEPQPKLPEKITKLFNR-DNIYFVGNPCTLNIHGFDTLIYHGRSFDDL : 220

      240           *           260           *           280           *
PHBN023 : VSFVPGLTHHKPGLPMVELLKMRLHAPTFGGKVPIAPDPE-DLLVIEEVPDLVQMGHV : 278
PfuORF2 : VGSVPGLTHHKPGLPMVELLKMRLHAPMFGGKVPIAPDPE-DLLVIEEVPDVVHMGHV : 278
DPD2_HUMAN : FRYSG---SMEDHLEILEWTLRVRHISPTAPDTLGCYPFYKTDPFIFPEC PHVYFCGNT : 275
MTH1405 : AMSVNGLSHERSDIMEELLEKRHLAPIYGERTPLASEIE-DHLVIDEVPHVLHTGHV : 274
AF1790 : ISKIPRLSYDEPQKVMEEELKRRHLSPIYGGRTPLAPERE-DYLVIEDVDPDILHCGHI : 271
MJ0702 : VGQIRAAASYENPVTIMKELIKRRLLCPTYGGRCPIAPEHK-DYLVIDRDIDILHTGHI : 277

      300           *           320           *
PHBN023 : HUYDTAVYRG-----VQLVNSATWQAQTEFQKMVNIVPTPGLVPIVD : 320
PfuORF2 : HUYDAVYRG-----VQLVNSATWQAQTEFQKMVNIVPTPAKVPVVD : 320
DPD2_HUMAN : PSFGSKIIRGPEDQTVLLVTVPDF-SATQTAQLVNLRL-SLACQPISF : 320
MTH1405 : HINAYKKYKG-----VHLINSGTFSQTEFQKIYNIVPTCGQVPVLN : 316
AF1790 : HTYGTGFYRG-----VFMVNSSTWQAQTEFQKKVNLNPMPGNVAVYR : 313
MJ0702 : HINGYGIYRG-----VVMVNSGTFSQTEFQKRMGISPTPAIVPIIN : 319

```



## Abbildung 17:

```

          *           20           *           40           *
DPOD_HUMAN : KVQSYEKEEDLLQAWSTFIRIMDPDVTGYNIQNFDLPYLISRAQTLKVQTFPFLGRV : 58
DPOL_ARCFU : EIILTGDERKIIISDFVKLIKSYPDIIVGYNQDAFDWPYLKRAERWNIPLD--VG-- : 54
MTH1208 : FVEVVEDERELLERFAEIVIDKKPDILVGYNSDNFDFPYITRRAAILGAELD--LG-- : 54
PHBT047 : YVEVVSSEMIKRLIRVIKEKDPDVIITYNGDNFDFPYLLKRAEKLGIKLL--LG-- : 54
MJ0885 : NIEVVKNEKELIKKIIETLKEY--DVIITYNGDNFDFPYLKARAKIYGIDIN--LG-- : 52

          60           *           80           *           100           *
DPOD_HUMAN : AGLCSNIRDSSFQSKQTGRDRTKVVSVMGRVQMDMLQVLLREYKLRSHTLNAVSPFHFL : 116
DPOL_ARCFU : -----RDGSNVVFRGG-----RPKITGRLNVDLYDIAMRISDIKIKKLENVAEFLG : 100
MTH1208 : -----WDGSKIRTMRG-FANATAIKGTVHVDLYPVMRRYMNLDRYTLERVYQELF : 104
PHBT047 : -----RDNSEPKMQMG-DSLAVEIKGRIHFDLFPVIRRTINLPTYTLEAVYEAIF : 104
MJ0885 : -----KDGEELKIKRGG-MEYRSYIPGRVHIDLYPISRLLKLT KYTLEDVVYNLF : 102

          120           *           140           *           160           *
DPOD_HUMAN : GEQKE-DVQHSIITDLQNGNDQTRRLAVYCLKDAYLPLRLLERLMVLVNAVEMARVT : 173
DPOL_ARCFU : TKIEIADIEAKDIYRYWSRGE--KEKVLNRYARQDAINTYLIK--ELLPMHYELSKMI : 154
MTH1208 : GEEKI-DLPGDRLWEYWDRLDEL-RDELFRYSLDDVVATHRIAE--KILPLNLELTRLV : 158
PHBT047 : GKPKE-KVYADEIAKAWETGEG-LERVAKYSMEDAKVTYELGR--EFFPMEAQLARLV : 158
MJ0885 : GIEKL-KIPHTKIVDYWANND---KTLEYSLODAKYTYKIGK--YFFPLEVMFSRIV : 154

          180           *           200           *           220           *
DPOD_HUMAN : GVPLSYLLSRGQOVVVSQLLRQAMHEGLLMPVVKSE-----GGEDYTGATVIEPLK : 225
DPOL_ARCFU : RLPVDDVTRMGRGKQVDWLLSEAKKIGEIAPNPPE-----HAESYEGAFVLEPER : 205
MTH1208 : GQPLFDISRMTGQQAWEFLVRKAYQYGELVPNKPSQSDFSRRGRRAVGGYVKEPEK : 216
PHBT047 : GQPVVDVSRSSGTNLVVEWFLLRKAYERNELAPNKPDEKEYERRLRESYEGGYVKEPEK : 216
MJ0885 : NQTPFEITRMSSGQMVEYLLMKRAFKENMIVPNKPDEEYRRRVLT TYEGGYVKEPEK : 212

          240
DPOD_HUMAN : GYYDVPIATLDFS : 238
DPOL_ARCFU : GLHEN-VACLDFA : 217
MTH1208 : GLHEN-IVQFDFR : 228
PHBT047 : GLWEG-IVSLDFR : 228
MJ0885 : GMFED-IISMDFR : 224

```

## Abbildung 18:

```

      *           20           *           40           *           60
AF1722 : DTIKGVKGMTSKTKIPERLEKGILRVKHGVFVKDGTARFDATDLPITHFKPAEIGVSVEKLR : 63
MJ1630 : GDVKCIKGMTSKQKIVEPLEKAILRAINEVYVFKDGTTRFDCTDVPVTHFKPNEINVTVEKLR : 63
PfuORF3 : DKLKGVMGMTSGWKIAEPLKGLLRKAKNEVYVFKDGTIRFDATDAPITHFRPREIGVSVEKLR : 63
MTH1536 : DEIKGVEGMISAEKFPPEPLEKGILRAKNDVYVFKDATIRHDSTDLP LTHFTPREVGVSVERLR : 63
PHBN021 : DKLKGVMGMTSGWKMPPEPLEKGLLRKAKNDVYVFKDGTIRFDATDAPITHFRPREIGVSVEKLR : 63

```

```

      *           80           *           100           *           120
AF1722 : ELGYERDYKGAELKNENQIVELKPQDVILPKSGAEYLLRVANFIDDLVVKFYKMEPFYNAKSV : 126
MJ1630 : ELGYDKDIYGNELVDGEQVVELKPQDVIIIPESCAEYFVKVANFIDDLLEKFKYKVERFYNVKKK : 126
PfuORF3 : ELGYTHDFEGKPLVSEDQIVELKPQDVILSKEAGKYLLRVARFVDDLLEKFGYGLPRFYNAEKM : 126
MTH1536 : ELGYTRDCYGDELEDEDQILELRVQDVVISDCADYLVRVANFVDDLLERFYDLERFYNVKTR : 126
PHBN021 : ELGYTHDFEGNPLVSEDQIVELKPQDIILSKEAGKYLLKVAKFVDDLLEKFGYGLPRFYNAEKM : 126

```

```

      *           140           *           160           *
AF1722 : EDLIGHLVIGLAPHTSAGVLGRIIGFSDVLAGYAHYPYFHAARR : 170
MJ1630 : EDLIGHLVIGMAPHTSAGMVGRIIGYTKANVGYAHYPYFHAARR : 170
PfuORF3 : EDLIGHLVIGLAPHTSAGIVGRIIGFVDALVGYAHYPYFHAARR : 170
MTH1536 : EDLVGHLIAGLAPHTSAAVLGRIIGFTGASACYAHYPYFHSARR : 170
PHBN021 : EDLIGHLVIGLAPHTSAGIVGRIIGFVDALVGYAHYPYFHAARR : 170

```

## Abbildung 19:

```

      *           20           *           40           *           60
DP3A_ECOLI : ELQVINQMGPFGYFLIVMEFIQWSKDNQVPGVPGRGSGAGSLVAYALKITDLDPLEFDLL : 60
DP3A_SALTY : ELQVINQMGPFGYFLIVMEFIQWSKDNQVPGVPGRGSGAGSLVAYALKITDLDPLEFDLL : 60
BB0579 : ELSVIIGMGPEGYFLIVWDFIKFAHNDIPVGAGRGSGAGSIVAYALRITDIDPLKYNLL : 60
DP3A_HELPI : EIEVITNMKFPGYMLIVWDFIRYAKEMGIPVGPGRGSAAGSLVAFALKITDIDPLKYDLL : 60
AA50 : ELEVINKMGFAGYFLIVQDFINWAKKNDIPVGPGRGSGAGSLVAYAIGITDVPDIKHGFL : 60

```

```

      *           80           *           100           *           120
DP3A_ECOLI : FERFLNPERVSMDFDVFDFCMEKRDQVIEHVADMYGRDAVSQIITFGTMAAKAVIRDVGR : 120
DP3A_SALTY : FERFLNPERVSMDFDVFDFCMEKRDQVIEHVADMYGRDAVSQIITFGTMAAKAVIRDVGR : 120
BB0579 : FERFLNPERISMDFDIDDFCFEGRDEIIEKYVTNKYGEDKVAQIITFGTLKPKAVVKDVAR : 120
DP3A_HELPI : FERFLNPERISMDFDIDDFCQRRRKEIIEYMIKEYGKYNVAQVITFNKMLAKGVIRDVAR : 120
AA50 : FERFLNPERVSMDFDIDVFDFCQDNREKVIEWVRNKYGHNDVQAQIITYNVMAKQTLRDVAR : 120

```

```

      *
DP3A_ECOLI : VLGHYPYGFVDRISKLIIP : 138
DP3A_SALTY : VLGHYPYGFVDRISKLVPP : 138
BB0579 : VLDIPFAESNELTKFIPD : 138
DP3A_HELPI : VLDMPYKEADDFAKLIPN : 138
AA50 : AMGLPYSTADKLAKLIPO : 138

```

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☒ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**

**THIS PAGE BLANK (USPTO)**